SEX ESTIMATION OF THE HUMAN SKELETON

SEX ESTIMATION OF THE HUMAN SKELETON

History, Methods, and Emerging Techniques

Edited by

ALEXANDRA R. KLALES

ELSEVIER

ACADEMIC PRESS
An imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, United Kingdom
525 B Street, Suite 1650, San Diego, CA 92101, United States
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom

© 2020 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-12-815767-1

For information on all Academic Press publications
visit our website at https://www.elsevier.com/books-and-journals

Publisher: Stacy Masucci
Acquisitions Editor: Elizabeth Brown
Editorial Project Manager: Ana Claudia A. Garcia
Production Project Manager: Selvaraj Raviraj
Cover Designer: Heather Garvin

Typeset by SPi Global, India

ELSEVIER
Book Aid International
Working together
to grow libraries in
developing countries
www.elsevier.com • www.bookaid.org

Contributors

Sarah E. Baumgarten
Anthropology, Washington University in St. Louis, St. Louis, MO, United States

Jonathan D. Bethard
Department of Anthropology, University of South Florida, Tampa, FL, United States

Soren Blau
Victorian Institute of Forensic Medicine/Department of Forensic Medicine, Monash University, Melbourne, VIC, Australia

Jaroslav Brůžek
Université de Bordeaux—CNRS—MCC, UMR 5199 PACEA, Pessac, France; Department of Anthropology and Human Genetics, Faculty of Science, Charles University, Prague, Czech Republic

Laura E. Cirillo
Department of Anthropology, University of Nevada-Reno, Reno, NV, United States

Stephanie J. Cole
Department of Anthropology, University of Nevada-Reno, Reno, NV, United States

Eugénia Cunha
Centre for Functional Ecology, Laboratory of Forensic Anthropology, Life Sciences Department, University of Coimbra, Coimbra, Portugal

Cassandra M. DeGaglia
Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC, United States

Itiel E. Dror
UCL Department of Security and Crime Science, Centre for the Forensic Sciences, London, United Kingdom

Evan M. Garofalo
Department of Basic Medical Sciences, University of Arizona College of Medicine—Phoenix, Phoenix, AZ, United States

Heather M. Garvin
Department of Anatomy, Des Moines University, Des Moines, IA, United States

Sara M. Getz
Department of Anthropology, Idaho State University, Pocatello, ID, United States

Pierre Guyomarc'h
Université Aix Marseille—CNRS—EFS, Marseille, France

Cortney N. Hulse
Department of Anthropology, University of Nevada-Reno, Reno, NV, United States

Richard L. Jantz
Department of Anthropology, University of Tennessee-Knoxville, Knoxville, TN, United States

Brittany Kenyon-Flatt
Anthropology, University of Buffalo, Buffalo, NY, United States

Alexandra R. Klales
Forensic Anthropology Program, Department of Sociology and Anthropology, Washburn University, Topeka, KS, United States

Natalie R. Langley
Department of Laboratory Medicine and Pathology, Division of Anatomic Pathology, Mayo Clinic, Scottsdale, AZ, United States

Holly Long
Forensic Anthropology Program, Department of Sociology and Anthropology, Washburn University, Topeka, KS, United States

Clare McFadden
School of Archaeology and Anthropology, The Australian National University, Canberra, ACT, Australia

Diana L. Messer
Department of Applied Forensic Sciences, Mercyhurst University, Erie, PA, United States

Ruth M. Morgan
UCL Department of Security and Crime Science, Centre for the Forensic Sciences, London, United Kingdom

Sherry Nakhaeizadeh
UCL Department of Security and Crime Science, Centre for the Forensic Sciences, London, United Kingdom

Stephen D. Ousley
Department of Anthropology, University of Tennessee-Knoxville, Knoxville, TN, United States

Nicholas V. Passalacqua
Anthropology and Sociology Department, Forensic Anthropology Program, Western Carolina University, Cullowhee, BC, United States

Marin A. Pilloud
Department of Anthropology, University of Nevada, Reno, NV, United States

Samantha K. Rowbotham
Victorian Institute of Forensic Medicine/Department of Forensic Medicine, Monash University, Melbourne, VIC, Australia

Frédéric Santos
Université de Bordeaux—CNRS—MCC, UMR 5199 PACEA, Pessac, France

G. Richard Scott
Department of Anthropology, University of Nevada, Reno, NV, United States

Michala K. Stock
Department of Exercise Science, High Point University, High Point, NC, United States

Kyra E. Stull
Department of Anthropology, University of Nevada-Reno, Reno, NV, United States; Department of Anatomy, University of Pretoria, Pretoria, South Africa

Richard M. Thomas
Federal Bureau of Investigation, Laboratory Division, Quantico, VA, United States

Douglas H. Ubelaker
Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC, United States

Caroline VanSickle
Department of Anatomy, A.T. Still University of Health Sciences, Kirksville, MO, United States

Cassidy Willsey
Forensic Anthropology Program, Department of Sociology and Anthropology, Washburn University, Topeka, KS, United States

About the contributors

Sarah E. Baumgarten, MS, MA[a]
Doctoral Candidate, Washington University at St. Louis
sbaumgarten@wustl.edu

Sarah E. Baumgarten is currently a PhD candidate at Washington University in St. Louis studying biological anthropology. Her current research focuses on morphological variation in hominin premolar tooth roots as a means of examining dietary adaptation. She is also a trained forensic anthropologist and currently works as a consultant for the Argentine Forensic Anthropology Team (EAAF). Her interests include biomechanics, evolutionary morphology, forensic anthropology, and its role in human rights. She holds an MA in anthropology from Washington University in St. Louis; MS in forensic and biological anthropology from Mercyhurst University; and BA in anthropology from Stony Brook University.

Jonathan D. Bethard, PhD, D-ABFA[a]
Assistant Professor, University of South Florida
jbethard@usf.edu

Jonathan D. Bethard maintains broad interests in bioarchaeology and forensic anthropology and pursues research questions in both fields. His current bioarchaeological project focuses on the medieval and early modern periods in Hungarian Transylvania. This work explores a number of questions related to better understanding the lives of people from this region through the biocultural lens of bioarchaeology. In addition, he has worked as an instructor for numerous courses in forensic anthropology with the International Criminal Investigative Training Assistance Program in Colombia and Algeria, as well as the International Committee of the Red Cross in Tbilisi, Georgia.

Soren Blau, PhD
Adjunct Associate Professor, Victorian Institute of Forensic Medicine/Department of Forensic Medicine, Monash University
soren.blau@vfm.org

Soren Blau is the manager of Human Identification and senior forensic anthropologist at the Victorian Institute of Forensic Medicine. She is an adjunct associate professor in the Department of Forensic Medicine at Monash University; Founding Fellow Faculty of Science, The Royal College of Pathologists of Australasia; and recipient of a Churchill Fellowship (2013). Soren is also currently the chair of the Forensic Anthropology

---

[a] Denotes first author.

Technical Working Group (National Institute of Forensic Sciences, Australia) and a member of the INTERPOL Disaster Victim Identification Pathology and Anthropology Sub-Working Group. Soren has published widely in the fields of forensic anthropology and archaeology.

Jaroslav Brůžek, PhD
Professor in the Department of Anthropology and Human Genetics, Faculty of Science, Prague, Czech Republic; and Emeritus Director of Research in the UMR5199 PACEA at the National Centre for Scientific Research (CNRS), University of Bordeaux, France
yaro@seznam.cz
Jaroslav Brůžek's research focuses on the methodology of reconstructing a biological profile from skeletal remains. He also studied human pelvic remains from the Upper Pleistocene to historical periods. His interest is sexual dimorphism in the context of bioarchaeology, and he is also co-author of publications on paleobiology of the early medieval population of Great Moravia. ORCID ID 0000-0002-2478-9662.

Laura E. Cirillo, MA
Doctoral Student, University of Nevada-Reno, Reno
lcirillo@nevada.unr.edu
Laura E. Cirillo is a doctoral student in biological anthropology at the University of Nevada-Reno, Reno. Her focus is on the growth and development of modern humans. Her research uses dentition for developing age estimation models and for the identification of developmental abnormalities through geometric morphometric analysis on 3D reconstructions. She has dedicated her time outside of research to forensic and bioarchaeological fieldwork, strengthening the potential of existing anatomical collections by introducing advanced analysis in a museum setting and collaborating with California tribes on NAGPRA repatriation efforts.

Stephanie J. Cole, MS[a]
Doctoral Candidate, University of Nevada-Reno, Reno
stephaniejcole@nevada.unr.edu
Stephanie J. Cole is a doctoral student in biological anthropology at the University of Nevada-Reno, Reno. Her primary research interests lie within the realm of forensic anthropology, specifically with regard to improving identification standards for both adults and subadults. Most of her research has focused on sex estimation. Recently, she assisted in the development of the adult morphological sex estimation database, MorphoPASSE. Her current research involves investigating the ontogeny of sex traits used in MorphoPASSE and the development of classification equations to assist in estimating sex in the young.

Eugénia Cunha, PhD
Professor, National Institute of Legal Medicine and Forensic Sciences; and University of Coimbra
cunhae@ci.uc.pt
Since August 2018, Eugénia Cunha has been the director of the National Institute of Legal Medicine and Forensic Sciences, South Delegation, Lisbon, and invited Full Professor at the University of Coimbra where she has been a teacher for 34 years. She is a forensic anthropologist; FASE-certified; a fellow of the American Academy of Forensic Sciences; a board member of FASE (where she was president 2009–16); and vice-president of Brazilian Association of Forensic Anthropology (ABRAF). Her research focuses on identification, namely age at death. ORCID ID 0000-0003-2998-371X.

Cassandra M. DeGaglia, BS
Research Assistant, National Museum of Natural History, Smithsonian Institution
DeGagliaC@si.edu
Cassandra M. DeGaglia graduated from the George Washington University in 2016 with a degree in biological anthropology. She has been at the National Museum of Natural History since 2014. In her time there, she has worked on projects concerning sexual dimorphism in modern human populations, methods for age estimation, paleodemographic reconstruction, and the impact of taphonomic changes on casework in forensic anthropology.

Itiel E. Dror, PhD
Senior Cognitive Neuroscience Researcher, University College London
i.dror@ucl.ac.uk
Itiel E. Dror (PhD Harvard) is a cognitive neuroscientist who is interested in the cognitive architecture that underpins expert decision-making. Dror's research, published in over 130 research articles, demonstrates how contextual information can influence judgments and decision-making of experts. He has worked in a variety of domains, from policing and aviation, to medical experts and bankers, showing that even hard-working and competent experts can make unintentional errors in evaluating data. Dr. Dror has worked with many agencies in various countries to minimize error and enhance decision-making.

Evan M. Garofalo, PhD[a]
Assistant Professor, University of Arizona College of Medicine-Phoenix
emgarofalo@email.arizona.edu
Evan M. Garofalo holds a BA from Tulane University, an MSc from Bradford University, and a PhD in functional anatomy and evolution from John's Hopkins University School

of Medicine. She is a founding member of the LGBTQQIAA interest group of the American Association of Physical Anthropologists (Queerbioanth). Dr. Garofalo is currently an assistant professor of anatomy at the University of Arizona College of Medicine-Phoenix where she is the faculty advisor for the LGBTQ in Medicine interest group. She, her wife, Anne, and two cats enjoy the dry heat in an historic Phoenix home.

Heather M. Garvin, PhD, D-ABFA[a]
Associate Professor of Anatomy, Des Moines University
Heather.Garvin-Elling@dmu.edu
Heather M. Garvin holds degrees from the University of Florida, Mercyhurst University, and John's Hopkins School of Medicine. She currently teaches gross anatomy to medical students at Des Moines University. She is a diplomat of the American Board of Forensic Anthropology, conducts forensic anthropology casework for the state of Iowa, and has been on the editorial board for the *Journal of Forensic Sciences* since 2015. She has over 30 publications and has done extensive research on human sexual dimorphism and factors contributing to population variation in dimorphism.

Sara M. Getz, PhD
Assistant Professor of Anthropology, Idaho State University
getzsara@isu.edu
Sara M. Getz holds a PhD in anthropology and BS in biological anthropology from Penn State University, as well as an MS in forensic and biological anthropology from Mercyhurst University. She specializes in the use of transition analysis and is broadly interested in the factors affecting variability in the human skeleton, how variation can be described and quantified effectively, and how it can contribute to investigations of individuals in forensic and archaeological contexts. She is also interested in research related to outdoor scene processing and forensic taphonomy. More information about Dr. Getz can be found on her website at saramgetz.com.

Pierre Guyomarc'h, PhD
Forensic Anthropologist, Université Aix Marseille
pierreguyo@gmail.com
Pierre Guyomarc'h graduated from the University of Bordeaux in France in 2008 and obtained a PhD on computerized facial approximation. He is actually working with the International Committee of the Red Cross on humanitarian forensic action since 2014. Pierre is affiliated to the University of Aix Marseille for research activities, and his interests focus on the creation and application of osteobiography methods for human identification in a forensic context. ORCID ID 0000-0002-9419-9270.

Cortney N. Hulse, MA
Doctoral Candidate, University of Nevada-Reno, Reno
chulse@nevada.unr.edu
Cortney N. Hulse is a PhD candidate at the University of Nevada-Reno, Reno, in biological and forensic anthropology. She specializes in trauma analysis, specifically she studies statistical methodologies in quantifying blunt-force trauma injuries. Her other interests include human rights and victim identification, as well as sex, age, and ancestry estimation. Aside from her research, she has worked internationally and domestically in bioarcheological fieldwork and works with local law enforcement and medical examiners on forensic recoveries and identifications.

Richard L. Jantz, PhD[a]
Professor Emeritus, University of Tennessee-Knoxville
rjantz@utk.edu
Richard L. Jantz is a biological anthropologist with research interests in human variation, forensic anthropology, and skeletal morphometrics with special reference to secular change in US populations over the past several centuries. He retired in 2010 and, since then, has enjoyed the opportunity to do uninterrupted research using the great databases at the University of Tennessee-Knoxville.

Brittany Kenyon-Flatt, MA
Doctoral Candidate, University at Buffalo
bkenyon@buffalo.edu
Brittany Kenyon-Flatt is a PhD candidate at the University at Buffalo, where she studies primate systematics and morphological variation using geometric morphometric methods. She holds an MA in anthropology from the University at Buffalo and a BA in anthropology from Ithaca College.

Natalie R. Langley, PhD, D-ABFA[a]
Associate Professor of Anatomy, Mayo Clinic Arizona
langley.natalie@mayo.edu
Natalie R. Langley is associate professor of anatomy in the Mayo Clinic Department of Laboratory Medicine and Pathology, Division of Anatomic Pathology at Mayo Clinic, Arizona. She received her PhD in anthropology in 2009 from the University of Tennessee in Knoxville. She became a board-certified forensic anthropologist in 2015 and is a member of the Board of Directors of the American Board of Forensic Anthropology. Her research interests include skeletal maturation, secular changes in skeletal biology, skeletal trauma, postmortem interval estimation, and clinical anatomy.

Holly Long, BS
Washburn University
halong177@gmail.com

Holly Long received her Bachelor of Science degree in anthropology (forensics concentration) and minor in biology at Washburn University, Topeka, Kansas, in Fall 2019. While at Washburn, she participated in active forensic casework under Dr. Klales; attended a field school in Roncesvalles, Spain; helped facilitate forensic short courses; and was the forensic anthropology lab assistant. She also conducted independent sex estimation research, which she presented at the 71st AAFS conference and submitted to *Forensic Anthropology*. She plans to pursue her master's degree in Fall 2020. She hopes to stay in forensics but also enjoys archaeological applications.

Clare McFadden, PhD
Lecturer, Biological Anthropology, Australian National University
clare.mcfadden@anu.edu.au

Clare McFadden has a Bachelor of Arts (Hons, 2015) and a PhD in biological anthropology (2019) from the Australian National University (ANU). She is an Associate Fellow of the Higher Education Academy. In 2019 she was appointed lecturer in biological anthropology in the School of Archaeology and Anthropology, ANU. Dr. McFadden leads a palaeodemography and palaeoepidemiology research group, exploring population dynamics and health globally from the Neolithic to the post-Medieval period. In parallel, she is further refining palaeodemographic methodologies relating to the estimation of sex, age, fertility, and various models of morbidity and mortality.

Diana L. Messer, PhD[a]
Mercyhurst University
dianamesser@gmail.com

Diana L. Messer received her doctoral degree in anatomy from The Ohio State University, a master's degree in forensic and biological anthropology from Mercyhurst University, and a bachelor's degree in biological anthropology from Southern Connecticut State University. She has worked on over 50 forensic cases and at bioarchaeological sites in Peru, Poland, and Oman. Her research focuses on radiographic time since injury estimation of healing fractures in children for application to child physical abuse cases. Diana currently works as an anthropologist in support of the Defense POW/MIA Accounting Agency Laboratory in Hawaii.

Ruth M. Morgan, PhD
Professor of Crime and Forensic Science, University College London (UCL)
ruth.morgan@ucl.ac.uk
Ruth M. Morgan [MA (Oxon), DPhil] is the founder and director of the UCL Centre for the Forensic Sciences, and professor of crime and forensic science at UCL. Her research group focuses on the interpretation of forensic science evidence. Professor Morgan has received the PW Allen Award from the Chartered Society of Forensic Sciences for excellent research three times in 2006, 2016, and 2018. Professor Morgan is a World Economic Forum (WEF) Young Scientist, a member of the WEF Global Future Council, and acted as the Specialist Adviser to the House of Lords Science and Technology Committee inquiry 2018–19.

Sherry Nakhaeizadeh, PhD[a]
Lecturer in Crime and Forensic Science, University College London (UCL)
sherry.nakhaeizadeh@ucl.ac.uk
Sherry Nakhaeizadeh is a lecturer in crime and forensic science at UCL Department of Security and Crime Science, Centre for the Forensic Sciences. Sherry's principle research interests include the effect of cognitive bias in forensic science and criminal investigations. She is particularly interested in judgment and decision-making in forensic anthropology and the approaches of evidence analysis in forensic science. Her research interest also looks into the use of eye-tracking technology in forensic decision-making, and how such technology could aid in an empirical approach to advancing the understanding of complex visual tasks in forensic anthropology, specifically, and forensic science, broadly.

Stephen Ousley, PhD
craniator@statsmachine.net
Stephen Ousley earned his PhD in biological anthropology at the University of Tennessee-Knoxville in 1997. With Richard Jantz, he co-authored FORDISC, a computer program that aids in the identification of unknown human remains using various statistical methods. From 1998 through 2007, Dr. Ousley was the director of the Repatriation Osteology Laboratory in the Repatriation Office of the National Museum of Natural History at the Smithsonian Institution. From 2007 to 2019, he was a professor at Mercyhurst University in Erie, Pennsylvania. He is now retired from full-time teaching in order to conduct research with his colleagues and friends.

Nicholas V. Passalacqua, PhD, D-ABFA
Associate Professor/Forensic Anthropology Program Coordinator, Anthropology and Sociology Department, Western Carolina University (WCU)
nvpassalacqua@wcu.edu

Dr. Passalacqua received his PhD in anthropology from Michigan State University in 2012 and was certified by the American Board of Forensic Anthropology in 2016. Prior to his arrival at WCU, he was a deploying forensic anthropologist with the Defense POW/MIA Accounting Agency in Oahu, Hawaii. Dr. Passalacqua is a co-founder and a current co-editor of *Forensic Anthropology*. He is also currently the vice president of the American Board of Forensic Anthropology, the chair of the Anthropology Consensus Body of the Academy Standards Board, and a member of the Anthropology subcommittee of the Organization of Scientific Area Committees.

Marin A. Pilloud, PhD, D-ABFA[a]
Assistant Professor, University of Nevada, Reno
mpilloud@unr.edu

Marin A. Pilloud is an assistant professor of anthropology at the University of Nevada, Reno. Prior to this appointment, she was a forensic anthropologist with the Central Identification Laboratory, Defense POW/MIA Accounting Agency. Her research focuses on the human skeleton and how it can inform our understanding of human behavior in archaeological contexts and also be used as part of the biological profile. She is a board-certified forensic anthropologist and a registered professional archaeologist. She has numerous publications in peer-reviewed journals and edited volumes, and has a co-edited volume as well as a co-authored book.

Samantha K. Rowbotham, PhD[a]
Forensic Anthropologist, Victorian Institute of Forensic Medicine/Department of Forensic Medicine, Monash University
Samantha.Rowbotham@vifm.org

Samantha K. Rowbotham is a forensic anthropologist with the Victorian Institute of Forensic Medicine (VIFM); an adjunct research fellow with the Department of Forensic Medicine, Monash University; and the first forensic anthropology trainee with the Royal College of Pathologists of Australasia (Faculty of Science). Samantha was a recipient of a VESKI Victoria Fellowship (2017) and the Australian Academy of Forensic Sciences "Oscar Rivers Schmalzbach" Research Fellowship (2019). Since 2009, she has been involved with archaeological excavations, osteology research, and bioarchaeology teaching projects in Asia, Europe, Central America, and Australia. Since 2017, she has undertaken forensic anthropology casework for the VIFM.

Frédéric Santos, MSc[a]
Engineer in Statistics, CNRS—Université de Bordeaux—Ministère de la Culture, France
frederic.santos@u-bordeaux.fr
Frédéric Santos' research activities focus on applied statistics and statistical modeling for past sciences, and especially biological anthropology through skeletal age and sex estimation. He is also teaching statistics and R programming at the University of Bordeaux. ORCID ID 0000-0003-1445-3871.

G. Richard Scott, PhD
Foundation Professor, University of Nevada, Reno
grscott@unr.edu
Before his initial appointment at the University of Nevada, Reno, in 2004, Richard Scott taught at the University of Alaska Fairbanks from 1973 to 1997, where he retired as emeritus professor. His research focuses on dental anthropology, with a special emphasis on tooth crown and root morphology. Geographic areas of interest include the American Southwest, Alaska, the North Atlantic, and Spain. He has co-authored three books on dental morphology and co-edited two other volumes. He has additional publications in peer-reviewed journals and edited books.

Michala K. Stock, PhD[a]
Assistant Professor, High Point University
mstock@highpoint.edu
Michala K. Stock is a biological anthropologist specializing in forensic anthropology, skeletal biology, and human anatomy. She is currently an assistant professor in the Department of Exercise Science at High Point University, and received her training at New York University (MA) and the University of Florida (PhD). As a graduate student, she conducted forensic anthropological casework at the C.A. Pound Human Identification Laboratory. Her research interests include investigating growth and development's effects on producing variation in the modern human skeleton.

Kyra E. Stull, PhD, D-ABFA
Assistant Professor, University of Nevada-Reno, Reno
kstull@unr.edu
Kyra E. Stull is an assistant professor in the Department of Anthropology at the University of Nevada-Reno, Reno. She is a biological anthropologist and a board-certified forensic anthropologist with specific interests in quantitative methods, growth and development, forensic anthropology, and modern human variation. She primarily uses advanced imaging techniques to answer research questions on subadult growth and has subsequently amassed a large database of CT and radiographic images of children from around the world.

Richard M. Thomas, PhD, D-ABFA[a]
Forensic Examiner, Federal Bureau of Investigation
rmthomas@fbi.gov

Richard M. Thomas received his BA in anthropology from Northwestern University and his MA and PhD in biological anthropology from the Pennsylvania State University. He is employed at the FBI Laboratory where his duties have included producing mitochondrial DNA sequence data from skeletal samples and conducting forensic anthropology examinations and research. He is a fellow of the Anthropology section of the AAFS and a diplomat of the American Board of Forensic Anthropology. He has served on the Scientific Working Group for Forensic Anthropology and on the Anthropology subcommittee of the Organization of Scientific Area Committees for the forensic sciences.

Douglas H. Ubelaker, PhD, D-ABFA[a]
Curator and Senior Scientist, Smithsonian Institution
ubelaked@si.edu

Douglas H. Ubelaker received his PhD from the University of Kansas in 1973. Since then, he has been employed at the Smithsonian Institution in Washington, DC. He has published extensively in forensic anthropology, focusing on both methods and case applications. He served as the 2011–12 President of the American Academy of Forensic Sciences and chairs the Humanitarian and Human Rights Resource Center of the AAFS. He is active in many international organizations and serves on numerous journal editorial boards.

Caroline VanSickle, PhD
Assistant Professor of Anatomy, A.T. Still University of Health Sciences
carolinevansickle@atsu.edu

Dr. Caroline VanSickle is an assistant professor of anatomy at A.T. Still University of Health Sciences, where she teaches gross anatomy to medical and dental students. She is an honorary affiliate of the Evolutionary Studies Institute at the University of Witwatersrand for her ongoing research on the skeletal anatomy of *Homo naledi*. Broadly, her research uses 3D imaging and comparative anatomy to study the evolution of sex differences in the hominin skeleton. She earned her PhD in anthropology from the University of Michigan, and was the inaugural Wittig Postdoc in Feminist Biology at the University of Wisconsin-Madison.

Cassidy Willsey, BS
Washburn University
cassidy.willsey@washburn.edu

Cassidy Willsey received her Bachelor of Science degree in anthropology (forensics concentration) and minor in biology at Washburn University, Topeka, Kansas, in Fall 2019. At Washburn, she assisted with forensic cases under the direction of Dr. Klales, hardening her decision to continue her pursuit in her education for forensic anthropology. After completing her undergraduate degree at Washburn, she plans to pursue a master's degree in anthropology.

About the editor

Alexandra R. Klales, PhD, D-ABFA (#123), is an Associate Professor in the Forensic Anthropology Program and Department of Sociology and Anthropology at Washburn University in Topeka, Kansas. She teaches courses in biological anthropology, forensic anthropology, and forensic archaeology. Dr. Klales earned her BA in anthropology from the University of Pittsburgh, Pittsburgh (2006), her MS in forensic and biological anthropology from Mercyhurst College (2009), and her PhD in anthropology from the University of Manitoba (2014). She is the founder and director of the Washburn University Forensic Anthropology Recovery Unit (WU-FARU), which assists local and regional law enforcement and coroners with active forensic casework in Kansas and Missouri. Dr. Klales is a Diplomate (#123) of the American Board of Forensic Anthropology and provides forensic anthropology consulting services for Frontier Forensics Midwest, the Kansas City Regional Mortuary Operational Response Group, the Missouri Disaster Response System, and the Kansas Bureau of Investigations. She is a member of the American Academy of Forensic Sciences (Anthropology Section) and the American Association of Physical Anthropologists. Her research includes improving biological profile methods, with a focus on sex estimation, and fatal fire/mass disaster recovery protocols. Dr. Klales has been awarded over $600,000 in federal grant funding, from which she created the MorphoPASSE sex estimation program (www.MorphoPASSE.com). She currently serves as Associate Editor for *Forensic Anthropology* and has published articles in the *American Journal of Physical Anthropology*; *Journal of Forensic Sciences*; and *Forensic Science International*, along with numerous book chapters. Dr. Klales also holds annual forensic anthropology short-course training seminars for first responders, law enforcement, coroner/medical examiners, anthropology students, and arson investigators at Washburn University.

Acknowledgments

I would first like to thank Samantha Rowbotham. This edited volume arose out of our discussions one cold Kansas day in February at a local coffee shop. I am thankful for her support and encouragement on this project along the way, her thoughts and feedback on the contents of the work, and for her chapter contribution to this volume.

Sincere thanks and gratitude go to each of the authors for their insightful contribution and commitment to advancing our understanding of skeletal sex estimation in humans. I would also like to thank the proposal reviewers who strengthened the overall focus of this volume and made many valuable recommendations: Eric Bartelink, Elizabeth DiGangi, Dennis Dirkmaat, Ericka L'Abbé, and Marin Pilloud.

I am grateful for L. Zachary DuBois's time and meaningful insight on the topic of transgender individuals over the course of several months. Those discussions deepened my understanding of the complex challenge of estimating and reporting sex/gender with regard to people who identify as transgender or who have medically transitioned.

Recognition and thanks go to the following colleagues (and friends) who have reviewed and edited various chapters or sections and provided feedback throughout this entire process: Heather Garvin, Nicholas Passalacqua, and Kyra Stull. Finally, thanks go to my family, especially my sister, for their support and tireless devotion to helping me succeed, even if that meant proof-reading the chapters multiple times and the dull task of checking references and page numbers.

Index

Note: Page numbers followed by *f* indicate figures and *t* indicate tables.

Introduction to sex estimation and this volume

Brief introduction to sex estimation

The biological profile or osteobiography consists of estimating an unknown individual's ancestry, sex, age, and stature based on their skeletal form. The purpose of estimating these demographic variables varies between bioarchaeology, forensic anthropology, and paleoanthropology (see Chapter 3 of this volume). Nevertheless, sex estimation is generally the second step in biological profile estimation and one of the most important, as many of the methods for stature and age estimation are sex-specific. While a crucial step of the biological profile, sex estimation cannot be performed until after the assessment of general age (adult vs. subadult) and estimation of ancestry. General age impacts which features and skeletal regions can or cannot be used (see Chapter 14 of this volume), and populations vary considerably in the levels of sexual dimorphism (see Chapter 17 of this volume). Sex estimation lacks some of the inherent difficulties found with the other profile parameters because the outcome is limited to only two options: male or female. However, despite having fewer options or "answers," sex estimation remains one of the more difficult aspects of biological profile assessment, especially with incomplete, fragmentary, and subadult remains or in populations with lower levels of sexual dimorphism.

Estimation of biological sex (see Chapter 4 of this volume) in skeletal biology is based on the premise that there are appreciable size and shape differences in the skeletal form of males and females within and between populations (i.e., sexual dimorphism). Humans are less sexually dimorphic than most of the other living primates, and secondary sex characteristics, which create the phenotypic differences between males and females, arise as a response to sexually dimorphic hormones (see Chapter 14 for a more detailed discussion). The degree of sexual dimorphism in any species is impacted by both extrinsic factors, like nutrition and stress, and intrinsic factors, such as hormone levels (see Moore, 2013 for a more detailed discussion). Essentially biological anthropologists are interpreting these sexually dimorphic features of the human skeleton. In any population, males will *on average* have more rugose or robust muscle attachment sites than females throughout the body. Females also exhibit appreciable differences in the form, positioning, and orientation of the pelvis related to the functional requirements of parturition and its effect on the bony pelvis (see Chapter 6 of this volume). These pelvic differences, in turn, result in male/female differences throughout the body, for example, in the carrying angle of the elbow and the Q-angle of the knee. In regard to size differences, males are, on average, anywhere from 8% to 10% larger (i.e., heavy, wider, or taller) than females (Rogers

& Mukherjee, 1992); therefore, males will tend to have larger skeletal measurements in comparison to females, especially in long bone length and joint size. Despite these differences, there will always be overlap between the sexes as there will be larger females and smaller males in any population. Considerable variation also exists within and between populations, and this must be considered when estimating sex of an unknown individual (see Chapter 17 of this volume).

Current methods used to estimate sex consist of either (1) qualitative traits sometimes referred to as nonmetric, morphological, morphoscopic, macromorphoscopic, anthroscopic traits or (2) quantitative measures known as metrics (see below for a more detailed discussion of appropriate terminology). The former consist of visually examining a particular feature, skeletal region, or trait and determining if it is robust/gracile or, in some cases, present/absent. The assumption is that the more gracile expression will have a greater frequency in females, while the more robust expressions will be more typical of males. In cases of presence or absence, the trait is question is presumed to be found in one sex rather than the other. The combination of these traits can be compiled for a majority rule or decision table approach (e.g., Burns, 2006; Phenice, 1969; Rogers & Saunders, 1994), or more appropriately incorporated into an established method with statistical measures of probability (e.g., Klales, Ousley, & Vollner, 2012; Rennie, 2018; Walker, 2005, 2008). The latter approach focuses on sexually dimorphic size (breadth and length) differences between the sexes, which can be captured metrically. One could argue for the addition of a third category or subcategory of metric methods called geometric morphometrics, which examines the interplay and differences between shape and size from a metric perspective (see Chapter 13 of this volume).

There remains debate as to which data type (qualitative vs. quantitative) and which specific methods we should be using to estimate sex (Garvin, 2012). A survey of practicing biological anthropologists indicated that the vast majority prefer using both qualitative and quantitative approaches to estimate sex; however, when only one or the other is utilized, qualitative methods were preferred nearly 2:1 (see Chapter 2 of this volume for complete survey results). Recent works have documented a perceived shift to quantitative methods due to their seemingly more objective nature (Christensen, Passalacqua, & Bartelink, 2019; Dirkmaat, Cabo, Ousley, & Symes, 2008; Moore, 2013), but one could also argue that many metric approaches suffer from reliability issues due to the difficulty in defining or locating landmarks (e.g., Type 2 and Type 3). Further criticisms of metric approaches suggest that there is greater population specificity; they are time-consuming and require specialized equipment; and they necessitate greater training. Criticisms of standard qualitative methods (majority rule or presence/absence of traits) include greater subjectivity, reliance on experience, and a lack of statistical rigor (Bruzek, 2002; Klales et al., 2012), yet they are quick and easy to apply. Stewart (1979) once argued why "waste time to measure traits that can be verified very quickly by the naked eye?" (Moore, 2013, p. 92). Based on survey results, we can clearly see that skeletal biologists are using both qualitative and quantitative approaches, neither of which is without its own benefits and inherent limitations. Several studies have demonstrated the correlation between metrics and morphological trait expression and suggest "there is no biological reason to favor either kind" (e.g., Cheverud et al., 1979, p. 196; Kenyhercz, Fredette, Klales, & Dirkmaat, 2012). It, therefore, stands to reason that we should be using as many possible sources of information to make an accurate estimate of sex, and often the qualitative and quantitative data will coincide.

Which particular method to use varies considerably from case to case, and will likely vary based on different contexts (e.g., bioarchaeological vs. forensic). The first primary limitation for appropriate method selection will always be which bones/features are available for analysis. For example, Waldron (1987) suggests that only about two-thirds of remains recovered in archaeological contexts contain the pubic bone, thereby negating the ability to use the highly accurate methods associated with this particular skeletal region. Because of differential preservation, virtually every single bone has been assessed quantitatively, qualitatively, or both for its potential for sex estimation due to the incomplete nature of skeletal remains that are often encountered in forensic, bioarchaeological, and paleoanthropological contexts. A second limiting factor is resource availability: Do you have the equipment necessary to use a particular method (e.g., calipers, manuals, etc.)? Do you have access to a computer, software, or internet connection? Are their financial constraints (e.g., purchasing software, digitizer, dental calipers, etc.)? In a perfect world, these would not be limiting factors, but, in reality, not all practitioners and laboratories have equal access to the materials or resources required to utilize certain methods. Lastly, determinations of method appropriateness should include a critical evaluation of a method's suitability for a specific population and to the tests of validity and reliability of the method. As Ousley always says, it is our responsibility to "do good science" (Klales et al., 2012, p. 106) of which *Daubert* reminds us in forensic settings, but which should also be the case in nonforensic contexts as well.

Terminology

Below is some clarification on the nuances of sex estimation terminology that are used throughout the chapters within this volume.

Sex assessment vs. sex determination vs. sex estimation

Sex *assessment* has been defined by Spradley and Jantz (2011, p. 290) as the use of "morphological traits with no estimable error rates, classification rates, or any associated statistics." This has been the historic approach to both sex and ancestry estimation in bioarchaeological and forensic contexts, whereby features or the *gestalt* were used to subjectively produce a sex assessment. The use of "assessment" for sex estimation is quite problematic based on the aforementioned definition. We know today that this historical assessment approach is not only invalid and unreliable but also lacks the scientific rigor required of our methodology; therefore, we should be moving away from both the practice of assessment and the use of the term assessment to refer to sex estimation practices. Recognizing the problematic issues with this terminology, the American Academy of Forensic Sciences' Academy Standards Board (ASB) Anthropology section recently (Fall 2019) corrected Standard 090 "Standard for Sex Assessment in Forensic Anthropology," based on feedback from the public commencing from January 2019.

Sex *determination*, on the other hand, implies levels of confidence approaching 100% accuracy and implies that sex can be treated as a known criterion (Gibbon, Paximadis, Štrkalj, Ruff, & Penny, 2009). The term *determination* itself is defined as "establishing something exactly" (*Oxford Dictionary*). At present, the only employable method with which to *determine* biological/chromosomal sex with near 100% accuracy is through DNA analyses, and even this method is not without its own caveats and limitations, such as false negatives (see Chapter 21 of this volume). Articles as recent as 2018/19 in the *Journal of Forensic Sciences*; *Forensic Science International*; *International Journal of Osteoarchaeology*; and the *American Journal of Physical Anthropology* include research methods on sex that are termed sex *determination*, rather than sex *estimation*, using methods other than DNA. Furthermore, current government agencies, such as the Defense POW/MIA Accounting Agency (DPAA), use sex *determination/assessment* within their standard operating procedures (SOPs) and case reports when estimating sex from skeletal parameters, even in circumstances when DNA is not utilized. Using the term "determination" in our case reports or site reports infers a level of confidence that simply cannot be obtained using currently available metric and morphological estimates of sex; therefore, the use of this terminology should be restricted to DNA analyses alone as it could confuse law enforcement, jurors, judges, coroners, the general public, and other agencies to which our reports are issued. Moore (2013, p. 92) perfectly summarizes this sentiment with her statement "until accuracy rates consistently reach 100% (which will likely never happen due to human variation), it is better to consider this endeavor estimation of sex."

Spradley and Jantz (2011, p. 290) define sex *estimation* as the use of "metric traits of the pelvis, skull, or any single bone or any combination of bones … because it provides an estimate in the form of an error rate or expected classification rate." Moore (2013) suggests the "current consensus in sexing research" focuses on metric methods, but I would argue that morphological methods are equally, if not more, popular (see Chapter 3 of this volume on practitioner preferences) due to their ease of use, broad applicability, and high agreement levels. Also, one could argue the modern morphological methods also include these estimates or statistical parameters (e.g., Klales, 2018; Klales et al., 2012; Walker, 2008) and, therefore, the term sex *estimation* should be expanded to include any estimation of sex with associated classification accuracy and error rates from skeletal parameters.

In the 21st century, we need to move away from using the term (and practice) of generating *assessments* and, instead, rely on *estimates* of sex (and other biological parameters) using valid and reliable methods (either morphological or metric). Our estimates should, in turn, include associated accuracy, probabilities, and error rates, and our methodological research at minimum should include these parameters, as well as tests of statistical assumptions (see Chapter 13 of this volume).

Nonmetric vs. morphological vs. qualitative methods vs. macromorphoscopic/morphoscopic

Colloquially, *nonmetric* simply means "not based on a standard of measurement," which is perhaps where the confusion arises with the usage of this term in skeletal biology (*Oxford Dictionary*). This layperson's definition simply means the assessment of anything that is not metric or measured as a continuous variable. However, in skeletal biology, nonmetric traits are often considered to be epigenetic (i.e., heritable) or quasi- or noncontinuous traits, which frequently vary by population group (Wilczak & Dudar, 2011) and, therefore, should be distinguished from the morphological skeletal variants used to estimate sex. A brief history of the term epigenetics suggests that its use to describe skeletal variants may also be problematic. Waddington (1968) originally defined the term as "the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being." This definition has since been refined to "the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence" (Wu & Morris, 2001). At present, the genetic mechanisms for many of the traits listed as nonmetric indicators of sex or ancestry are not well understood, and many of these traits may be less controlled by the epigenome and more so by environmental factors. The standardized skeletal documentation software Osteoware includes 62 nonmetric traits, some of which are likely not true nonmetric traits and are rather simple morphological traits (Wilczak & Dudar, 2011). Examples of true nonmetric skeletal traits include cranial ossicles, tori, septal apertures, third trochanters, and metopism. These traits have all been shown to exhibit a high heritability coefficient and are more strongly controlled by genetic, rather than environmental, factors (Sjøvold, 1984). While some of these traits tend to have higher frequencies in one sex or another, they are not typically utilized in a capacity for sex estimation.

Qualitative refers to describing the quality of something based on size, appearance, or value rather than on its quantity, while morphology is the form (size and shape) or structure of things (*Oxford Dictionary*). *Morphological* is most appropriate when we discuss sexually dimorphic skeletal features, because it is more broadly defined as form and structure and excludes measurable size differences and values. So, while we are using qualitative data, our methods are morphological. In the case of skeletal analyses, we are concerned with the internal anatomy of a person as reflected by the form of their skeletal features. Many of these features occur on a continuous gradation, for example, mastoid size or brow-ridge shape, which are often difficult to capture metrically. In some methods, these traits are scored on gradients from gracile to robust that can mimic quasicontinuous scoring found with nonmetric traits; however, they are not truly epigenetic, and these methods often group multiple features into a single scoring scale. For example, scoring the mental eminence in Walker (2008) requires assessment of tubercle presence/absence, width of the eminence, and projection of the chin (Lewis & Garvin, 2016). Likewise, in Klales et al. (2012), the ligamentous attachment orientation and angle are considered with overall bone shape for the evaluation of the ventral arc. In reality, multiple shape features are being grouped into scores based on overall morphology rather than scoring a single semicontinuous trait. Because of this, I would argue that we should be moving away from the use of the terms *nonmetric* or *qualitative* when we are referring to morphological traits used to differentiate the sexes.

Recent textbooks in forensic anthropology (cf., Christensen, Passalacqua, & Bartelink, 2014; Dirkmaat, 2012) suggest the use of terms *morphoscopic/macromosphoscopic* to describe traits that are scored based on presence/absence, degree of expression, or overall morphology (Christensen et al., 2014 definition). The term macromorphoscopic was first introduced in the literature by Ousley and Hefner (2005) to specifically refer to skeletal traits used in ancestry estimation, which was later shortened to morphoscopic in subsequent publications (e.g., Hefner, 2009; Hefner & Ousley, 2014; Hefner, Ousley, & Dirkmaat, 2012). They define the term as traits that "are quasicontinuous variables of the cranium that can be reflected as soft-tissue differences in the living" that are expressly used to "assess the ancestry of a single individual for the purpose of identification" (Hefner et al., 2012, p. 295). The authors prefer to restrict using this terminology to ancestry estimation of the skull only; however, the term has since been appropriated to describe all morphological traits of the skeleton (cf., Byrnes, Kenyhercz, & Berg, 2017; Christensen et al., 2014). If we look back to the first use of the term morphoscopic in the 18th century, it comes from the discipline of geology and refers to the observation of shapes on a microscopic scale, originally applied to sediments. Given that we are examining skeletal traits on a macroscopic rather than microscopic scale in sex estimation, the application of the term morphological is more appropriate than morphoscopic or macromorphoscopic, which should be restricted to the ancestry traits introduced by Ousley and Hefner (2005).

Sex vs. gender

See Chapter 4 of this volume for a detailed discussion of the differences and relationships between sex and gender. Broadly speaking, in skeletal sex estimation, we should move away from the use of gender and the nouns men and women, which infer two of several possible gender/identities rather than biological sex. Instead, research and reports in skeletal biology should refer to biological sex differences and use the nouns male and female. For a broader discussion of the inherent problems with this binary classification, see Chapter 4. We should never attempt to estimate a person's gender—which may have been fluid and dynamic in life—based on their skeletal morphology; however, some would argue that, perhaps, we can speak to gender based on the personal effects associated with remains, as is often the case with bioarchaeological interpretations. Recently, Kincer and Tallman (2019) administered a survey entitled "Transgender Knowledge in Forensic Anthropology" to the American Academy of Forensic Sciences listserv in an attempt to "better understand forensic anthropologists' knowledge of and experience with identifying transgender individuals in an effort to increase research in this area" (text from Survey Introduction). The survey asked practitioners several questions related to biological sex and gender—for example: Is human biological sex binary? Do you believe that forensic anthropologists should report on the gender of an individual in forensic anthropological casework? Upon identifying a transgender individual in forensic casework, how would you report the individual's sex? Results from their study are our first step in understanding current practices and research needs in this area. At present, research is very limited in this regard, and there are no discipline-wide guidelines on how to approach interpretations of gender in relation to biological sex based on skeletal morphology.

Masculine/feminine expression vs. robust/gracile expression

The terms masculine and feminine refer to attributes, behaviors, and roles associated with specific genders based on socially constructed norms, which can vary by culture and group (Shehan, 2018). As such, we cannot discern these from an individual's skeleton; thus, the use of terms masculine/feminine expression is incorrect. Walker (2008) and Klales et al. (2012) push for an objective scoring system from 1 to 5 that covers the range of human variation, rather than scoring skeletal features as the "masculine" or "feminine" expression or probable/definite male or female. Furthermore, the median value on these ordinal scales cannot be assumed to be the cut-off score between males and females for classification. We need to move away from the idea of scores 1/2 or −2/−1 being female or probable female, 3/0 as indeterminate, and 4/5 or +1/+2 being male or probable male, as originally presented in works such as Acsádi and Nemeskéri (1970) and Buikstra and Ubelaker (1994). There are always going to be both males and females present on the entire scale of human variation. The purpose of the scale is to simplify and objectify scoring with no presumptions of sex. Instead, we are scoring the robusticity/gracility or size/shape of a specific trait, and the statistics or method will calculate sex probability for the observer. This sentiment is explicitly stated in Klales et al. (2012): traits are "scored on an ordinal scale from one to five without any assumption of 'maleness' (masculinity) or 'femaleness' (femininity)." Unfortunately, this work and that of Walker (2008) continue to be erroneously cited and incorrectly summarized (e.g., Jones, 2014; Rennie, 2018), which further perpetuates the incorrect use of this terminology and assumptive practices.

Bone terminology

Bone terminology in this work has been kept consistent with those used in White, Black, and Folkens (2012) with the exception of the innominate. The use of the term skull is restricted to instances in which the crania and mandible are being discussed. The use of the term pelvis is restricted to instances including multiple bones of the entire structure (innominates, sacrum, coccyx). In this work, the hip bone is referred to as the innominate rather than the Latin *os coxae* (singular) or *ossa coxae* (plural) as presented in White et al. (2012), because none of the other bones of the body are referred to by their Latin names. It remains difficult to trace the origin of these terms, but it should be noted that *os coxa* and *os innominatum/ossa innominata* (Todd, 1920) are not anatomical terms or the proper Latin form of "hip bone."

Introduction to this book

Goals of this work

Sex estimation remains one of the most vital components of the biological profile or osteobiography, as many methods for estimating the remaining parameters are sex-dependent. At present, there are no field-wide best practices and guidelines within bioarchaeology, forensic anthropology, or paleoanthropology that indicate which methods should, or should not, be used to estimate the biological profile parameters, including sex. Instead, individual practitioners and laboratories base the final profile estimation on a combination of preferred methods and personal experience. This volume provides analysts with a starting point for research and practice on sex estimation in order to assist with the identification and analysis of human remains. It contains a comprehensive collection of the latest scientific research, theory, and methods in sexing human remains. Case studies are presented where relevant to highlight methodological application to real cases. This volume is designed primarily for professionals in fields who utilize skeletal biology for the purposes of sex estimation, including bioarchaeology, forensic anthropology, and paleopathology. However, this book can also complement advanced undergraduate and graduate student coursework and research in skeletal biology. This work also attempts to standardize some of the terminology and methods we are using with sex estimation (see above).

Format of this book

The book is divided into three main sections. The first section provides an overall introduction to sex estimation beginning with a comprehensive history and practitioner preferences. Then, the overlapping needs and clear differences between the fields utilizing skeletal biology for sex estimation are covered along with a discussion of biological sex and gender. Lastly, the impact of correct or incorrect sex estimation on the other biological profile parameters is explored. The second section addresses the main methodologies and skeletal regions used to estimate sex, including metric and morphological approaches, statistical applications, and software/computer programs. Each chapter provides a review of older techniques and emphasizes the latest research and methodological improvements. Case studies are presented where appropriate to highlight methodology. Finally, the third section addresses current considerations and future directions for sex estimation in forensic and bioarchaeological contexts. Topics covered in the last section include bias secular change, DNA analyses, medical imaging, population variation, and the impact of skeletal asymmetry on sex estimation.

The authors included in this volume vary from well-established researchers and professionals to up-and-coming graduate students who are participating in the latest research in sex estimation methodology. An attempt was also made to include perspectives from both the United States and internationally, as well as to include both bioarchaeologists and forensic anthropologists.

Conclusion

My hope with this work is to put sex estimation on equal footing with the other biological profile parameters that have been extensively explored and consolidated in the last 10 years (cf., Pilloud & Hefner's, 2016 *Biological Distance Analyses: Forensic and Bioarchaeological Perspectives*, or Latham & Finnegan's, 2010 *Age Estimation of the Human Skeleton*). While several works have touched on sex estimation and done an excellent job of summarizing it within larger edited volumes (e.g., Cabo, Brewster, & Azpiazu, 2012; Garvin, 2012; Moore, 2013), relegating it to a single chapter or two limits our complete and full understanding of sex estimation. The goal of this volume is to present a truly comprehensive representation of the current state of sex estimation while also detailing the history and how we got to this point.

Alexandra R. Klales
Forensic Anthropology Program, Department of Sociology and Anthropology,
Washburn University, Topeka, KS, United States

References

Acsádi, G., & Nemeskéri, J. (1970). *History of human life span and mortality*. Budapest: Akadémiai Kiadó.

Bruzek, J. (2002). A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology*, *117*, 157–168.

Buikstra, J. E., & Ubelaker, D. H. (1994). *Standards for data collection from human skeletal remains: Proceedings of a seminar at the Field Museum of Natural History, organized by Jonathan Haas* (3rd ed.). Fayetteville, AR: Arkansas Archeological Survey.

Burns, K. R. (2006). *Forensic anthropology training manual* (2nd ed.). New York: Pearson.

Byrnes, J. F., Kenyhercz, M. W., & Berg, G. E. (2017). Examining interobserver reliability of metric and morphoscopic characteristics of the mandible. *Journal of Forensic Sciences*, *62*, 981–985.

Cabo, L. L., Brewster, C. P., & Azpiazu, J. L. (2012). Sexual dimorphism: Interpreting sex markers. In D. C. Dirkmaat (Ed.), *A companion to forensic anthropology*. Wiley-Blackwell: West Sussex, UK.

Cheverud, J. M., Buikstra, J. E., & Twichell, E. (1979). Relationships between non-metric skeletal traits and cranial size and shape. *American Journal of Physical Anthropology, 50*, 191–198.

Christensen, A. M., Passalacqua, N. V., & Bartelink, E. J. (2014). *Current methods in forensic anthropology* (1st ed.). United Kingdom: Elsevier.

Christensen, A. M., Passalacqua, N. V., & Bartelink, E. J. (2019). *Current methods in forensic anthropology* (2nd ed.). United Kingdom: Elsevier.

Dirkmaat, D. C. (2012). *A companion to forensic anthropology*. West Sussex, UK: Wiley-Blackwell.

Dirkmaat, D. C., Cabo, L. L., Ousley, S. D., & Symes, S. A. (2008). New perspectives in forensic anthropology. *American Journal of Physical Anthropology*, *137*, 33–52.

Garvin, H. M. (2012). Adult sex determination: Methods and applications. In D. C. Dirkmaat (Ed.), *A companion to forensic anthropology*. Wiley-Blackwell: West Sussex, UK.

Gibbon, V., Paximadis, M., Štrkalj, G., Ruff, P., & Penny, C. (2009). Novel methods of molecular sex identification from skeletal tissue using the amelogenin gene. *Forensic Science International*, *3*, 74–79.

Hefner, J. T. (2009). Cranial nonmetric variation and estimating ancestry. *Journal of Forensic Sciences*, *54*(5), 985–995.

Hefner, J. T., & Ousley, S. D. (2014). Statistical classification methods for estimating ancestry using morphoscopic traits. *Journal of Forensic Sciences*, *59*(4), 883–890.

Hefner, J. T., Ousley, S. D., & Dirkmaat, D. C. (2012). Morphoscopic traits and assessment of ancestry. In D. C. Dirkmaat (Ed.), *A companion to forensic anthropology*. Wiley-Blackwell: West Sussex, UK.

Jones, G. (2014). *Not a yes or no question: Critical perspectives on sex and gender in forensic anthropology* [unpublished master's thesis] Ontario, Canada: University of Windsor.

Kenyhercz, M. W., Fredette, S. M., Klales, A. R., & Dirkmaat, D. C. (2012). Metric and non-metric assessment of sex: Accuracy, correlation and corroboration. *Proceedings of the American Academy of Forensic Sciences*, *18*, 406–407.

Kincer, C., & Tallman, S. (2019). *Transgender knowledge in forensic anthropology survey*. Qualtrics, Boston University School of Medicine.

Klales, A. R. (2018). *MorphoPASSE: The morphological pelvis and skull sex estimation database. Version 1.0.* Washburn University: Topeka, KS.

Klales, A. R., Ousley, S. D., & Vollner, J. M. (2012). A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *American Journal of Physical Anthropology*, *149*(1), 104–114.

Latham, K., & Finnegan, M. (2010). *Age estimation of the human skeleton*. Springfield, IL: Charles C. Thomas.

Lewis, C., & Garvin, H. M. (2016). Reliability of the Walker cranial nonmetric method and implications for sex estimation. *Journal of Forensic Sciences*, *61*(3), 743–751.

Moore, M. K. (2013). Sex estimation and assessment. In L. DiGangi & M. Moore (Eds.), *Research methods in skeletal biology* (pp. 91–116). New York: Academic Press.

Ousley, S. D., & Hefner, J. T. (2005). *The statistical determination of ancestry*. *Proceedings of the 57th annual meeting of the American Academy of Forensic Sciences, New Orleans, LA*. Colorado Springs, CO: American Academy of Forensic Sciences.

Phenice, T. W. (1969). A newly developed visual method for sexing the *os pubis*. *American Journal of Physical Anthropology*, *30*, 297–302.

Pilloud, M., & Hefner, J. T. (2016). *Biological distance analyses: Forensic and bioarchaeological perspectives*. New York: Academic Press.

Rennie, S. R. (2018). *Summary sex: A multivariate approach to sex estimation from the human pelvis* [unpublished doctoral dissertation]. Liverpool, England, United Kingdom: Liverpool John Moores University.

Rogers, A. R., & Mukherjee, A. (1992). Quantitative genetics of sexual dimorphism in human body size. *Evolution*, *46*(1), 226–234.

Rogers, T. L., & Saunders, S. (1994). Accuracy of sex determination using morphological traits of the human pelvis. *Journal of Forensic Sciences*, *39*, 1047–1056.

Shehan, C. L. (2018). *Gale researcher guide for: The continuing significance of gender* (pp. 1–5). Cengage Learning: Gale.

Sjøvold, T. (1984). A Report on the heritability of some cranial measurements and non-metric traits. *Multivariate Statistical Methods in Physical Anthropology*, 223–246.

Spradley, M. K., & Jantz, R. L. (2011). Sex estimation in forensic anthropology: Skull versus postcranial elements. *Journal of Forensic Sciences*, *56*(2), 289–296.

Stewart, T. D. (1979). *Essentials of forensic anthropology*. Springfield, IL: Charles C. Thomas.

Todd, T. W. (1920). Age changes in the pubic bone. I. The male white pubis. *American Journal of Physical Anthropology, 3*, 285–334.

Waddington, C. H. (1968). Towards a theoretical biology (pp. 1–32). Edinburgh, Scotland: Edinburgh University Press. The Basic Ideas of Biology.

Waldron, T. (1987). The relative survival of the human skeleton: Implications for paleopathology. In A. N. Boddington & R. C. Janaway (Eds.), *Death, decay and reconstruction* (pp. 55–64). Manchester: Manchester University Press.

Walker, P. L. (2005). Greater sciatic notch morphology: Sex, age, and population differences. *American Journal of Physical Anthropology*, *127*, 385–391.

Walker, P. L. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, *136*, 39–50.

White, T. M., Black, M., & Folkens, P. (2012). *Human osteology* (3rd ed.). New York: Elsevier.

Wilczak, C. A., & Dudar, J. C. (Eds.), (2011). *Vol.I Osteoware*^{TM} *Software Manual*. Washington, DC: Smithsonian Institution.

Wu, C. T., & Morris, J. R. (2001). Genes, genetics, and epigenetics: A correspondence. *Science*, *293*(5532), 1103–1105.

CHAPTER 1

A history of sex estimation of human skeletal remains

Alexandra R. Klales, Holly Long, Cassidy Willsey
Forensic Anthropology Program, Department of Sociology and Anthropology, Washburn University, Topeka, KS, United States

Introduction

Our understanding of the skeletal differences between males and females grew from an appreciation of human anatomy and later with a specific emphasis on human osteology. The historical review below focuses on advancements in skeletal anatomy and biology through the lens of sex estimation. Our history begins with religious, biblical, and early philosophical explanations and then moves to more scientific approaches with anatomy, human osteology, forensic applications, and then finally research and methodology development. This comprehensive history extends up to the mid-20th century (around the 1950s), as much of what has been done since then has already been extensively covered elsewhere (cf. Moore, 2013) and is also covered in part within the history sections of chapters in this volume.

History of human skeletal sex estimation

Differences between the sexes have "puzzled humanity since ancient times- not only within a scientific context, but also within a social context" (Stévant, Papaioannou, & Nef, 2018, p. 8). The review presented below speaks about sex differences from a binary perspective (male and female) based on historical works, but the authors recognize that both gender and biological sex exist along a spectrum (see Chapter 4 in this volume for a more in-depth discussion of sex, gender, and terminology from both a sociocultural and a biological perspective).

Prior to our understanding of human anatomy and osteology, many religious and philosophical explanations existed for the differences between females and males. For example, Greek philosopher Aristotle (335 BC) based sex differences on earthly elements: males had more fire and females had more water and that temperature during intercourse determined the sex of the offspring (Gardiner & Swain, 2015; Stévant et al., 2018). This notion of environmental sex determination in humans remained popular and persisted until the turn of the 20th century (Gardiner & Swain, 2015). Christian biblical accounts

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00001-8

explain the "design" of females when God created Eve by taking a rib from Adam, leading to the long-held erroneous belief that males have one fewer rib than females. Even Earnest Hooton, one of the most influential biological anthropologists, theorized that females had fewer ribs to accommodate an expanded uterus during pregnancy (Hooton, 1946). These beliefs have carried through Western medicine for years and are still sometimes perpetuated today despite scientific inquiry and affirmation proving otherwise.

Human anatomy

The first recorded account of what we would today call human internal anatomy comes from early ancient Egyptian texts (3000–2500 BC), quite unsurprisingly tied to both natural and artificial mummification, the latter of which required extensive anatomical knowledge (Persaud, 2014). There was a distinction made in the burial practices between the sexes in ancient Egypt, but there is no mention of known skeletal differences. Later, more evidence of anatomical studies appeared in India beginning in 1500 BC with the *Vedas* text (Zysk, 1986). An understanding of skeletal anatomy was rather limited likely due, in part, to the restrictions and cultural stigma associated with anatomical dissections at this time. The first record of the study of skeletal elements occurred around 600 BC—thousands of years after the first recorded study of human anatomy—by a physician named Susutra who dissected children in an attempt to better understand the body. At the time, any deceased individual over the age of two was burned following Hindu practice; therefore, he was restricted to young subadults. His records describe the body as having 300 bones, likely due to studying children with unfused epiphyses (Persaud, 2014; Zysk, 1986). Later Hippocrates (420 BC), known for his contribution to medicine and the creation of the Hippocratic Oath, wrote several treatises on medicine, including knowledge of human skeletal form. In his chapter "On Injuries of the Head," he noted human variation in sutures (Adams, 1886). Osteological knowledge was extremely limited at this time, which in turn limited the knowledge of skeletal differences between the sexes.

Human osteology

Galen's *De Ossibus ad Tirones* (translated as *On Bones for Beginners*) (AD 180) was the first published work dedicated entirely to human osteology. He noted differences between males and females in the neck of the femur, indicating females angled less than males (Singer, 1956). He also noted a harder, more articulated skeleton in fetuses of males versus females, with males developing more quickly (Connell, 2000). Although some knowledge of osteology and perceived sex differences appeared in medical and anatomical texts, it was not until the 13th century that a record of osteological applications to differentiating the sexes from the skeleton was found in *The Washing Away of Wrongs* by Song Tz'u (1247). This text is widely considered the earliest record of forensic sciences.

The work includes crude methods for differentiating males and females based on coloration of the remains, with male bones being white and female bones being darker (Tz'u 1247; translated by McKnight, 1981). Tz'u also believed there were sex differences in the number of cranial bones with males having eight, whereas females had only six due to a lack of "a vertical [suture] running down the hairline in the back" (McKnight, 1981, p. 34). Such a divergence in bone quantity was attributed to the importance of numerology in ancient China (McKnight, 1981). Though most of these noted sex differences have since been disproven due to the great variety of factors that can cause discoloration to the bones or fused sutures, these early works were revolutionary in their specific focus on human osteology.

Describing and understanding sex differences

We begin to see the first studies of sex differences on the skeletal form during the 16th century; however, anatomists of this time said very little on the subject as their primary focus was not on questions of sex differences (Schiebinger, 1986). The anatomists Vesalius (1543), in *De Humani Corporis Fabrica* (*On the Fabric of the Human Body*), and Bauhin and de Bry (1605), in *Theatrum Anatomicum*, were among the first to distinctly illustrate and/or describe the female skeleton in comparison to the male skeleton, implying that there were enough differences between the two to warrant separate representations. In each case, females were contrasted against the male form, likely indicating a level of perceived inferiority of women that was common at the time. Schiebinger (1986, p. 42) suggests that "defining sex differences in every bone … became a research priority" in Europe from 1730 to 1790. Notable works include Albinus (1749) in *Table of the Skeleton and Muscles of the Human Body*, Cowper (1737), Moreau (1750), Ackermann (1788), and von Soemmerring (1796) in *Tabula Sceleti Feminini Juncta Descriptione* (*Descriptions of the Female Skeleton*). Cowper's (1737) *The Anatomy of Humane Bodies* describes the different skeletal proportions between males and females primarily based on shoulder and hip breadth. Moreau (1750) published *A Medical Question: Whether Apart from Genitalia There Is a Difference Between the Sexes*, and Ackermann (1788) described skeletal sex differences, but also lamented that current understandings of sex differences were inadequate and arbitrary (Schiebinger, 1986). For a much more detailed and nuanced discussion of the history of sex differentiation research during this specific time period, see Schiebinger (1986). Although differences were noted through the 16th–17th centuries, no specific methods were developed for *estimating* sex from skeletal remains.

Shortly thereafter, rudimentary methods of sex estimation were being developed in the 1800s using cranial measurements as a byproduct of race studies. These were primarily focused on classifying groups, estimating cranial size, and then linking these variables to intelligence. In the early 1830s in Philadelphia, Samuel Morton examined cranial capacity between different populations based on a large collection of skulls. Morton assigned sex to each crania by overall size based on the notion that males were larger (Lewis et al., 2011). Later, in the 1840s in Sweden, Retzius developed the cephalic index (maximum cranial width/maximum cranial length × 100) to classify races, yet the index showed sex differences with females averaging a higher index than males (Hrdlička, 1919; Woo, Jung, & Tansatit, 2018). In 1873, Dureau, a French biological anthropologist, reviewed the methods that were currently available for sex estimation. He concluded that the sexes could be distinguished 90% of the time based on "accentuated character" states that demonstrate skeletal differences between the sexes (Giles & Elliot, 1963). French anatomist Broca also suggested both morphological and metric sex differences in cranial form in *Instructions Craniologiques et Craniometriques* (1875). Works at this time were rather limited and did not focus specifically on sex estimation.

Forensic applications

In the United States, early application of sex estimation to forensic casework began with the highly publicized murder of Dr. Parkman, a Harvard University professor, whose remains were discovered burned and scattered within a furnace, a privy vault, and a tea chest in 1849. Dr. Jeffries Wyman, a medical doctor from Harvard, used his knowledge of anatomy to determine that the remains were from the same male individual (Ubelaker, 2018). This landmark case can be found in virtually every introductory text on forensic anthropology that includes a history of the discipline. Thomas Dwight, who is widely considered by some to be the father of forensic anthropology and who eventually took over the Parkman Professorship of Anatomy at Harvard University, also consulted in a widely unknown forensic case referred to as the "headless skeleton mystery" (Harrington, 1911; Portland Transcript, 1874). Dwight testified in the trial of James Lowell, who stood accused of murdering his wife, and determined that the remains were in fact human and female (Harrington, 1911; Portland Transcript, 1874): "the sex of the skeleton is unquestionably female, as demonstrated by the general lightness of the bones, and particularly by the size and shape of the pelvis, not to speak of the character of the garments in which it was found encased" (Lowell, Plaisted, Maine Supreme Judicial Court, 1875, p. 201). His experience with this case led to his well-known publication of the *Identification of the Human Skeleton: A Medico-Legal Study* (Dwight, 1878). In his essay presentation to the Massachusetts Medical Society, Dwight details how the human skeleton can be used to answer a number of questions related to the biological profile, including sex, and to the postmortem interval (Harrington, 1911). Within this work and subsequent publications, Dwight focused on sex differences in bone measurements (e.g., sternal length and joint dimensions) and fusion patterns (Dwight, 1878, 1894; Hrdlička, 1919). George Dorsey, having studied under Dwight, also recognized that sex differences could be detected throughout the body (Dorsey, 1897). Specifically, he determined the "head of the humerus is a better indicator of sex than the head of the femur" (Stewart, 1979). In 1897, Dorsey was instrumental in analyzing the skeletal remains from the Luetgert sausage factory murder. Police found clothing, two rings, and some bones in a vat of a sausage factory and suspected those remains to belong to the sausage maker's wife Louise. Dorsey consulted on the case and determined the metatarsal bone, toe phalanx, sesamoid bone, and rib head were from a human female (Snow, 1982). In 1899, Dorsey gave a lecture on "The Skeleton in Medico-Legal Anatomy" that presented his knowledge and research as beneficial to more applications than simple anatomical study (Stewart, 1979). Shortly thereafter, Derry (1909) noted the link between the preauricular sulcus with parity and its potential as an indicator of sex. Much of his work was based on predecessors who focused on pelvic differences related to race rather than sex (Ubelaker & De La Paz, 2012).

In the first half of the 19th century, notable anthropologists Krogman, Todd, Hrdlička, Hooton, and Stewart began regularly consulting on forensic casework that often required estimation of sex for identification purposes. Likely as a result of this casework, many of these individuals began publishing on sex differences in the skeleton. In 1919, Hrdlička published an article discussing the differences in cranial features between male and female skeletons, based on measurements of the vault. He also discussed shape differences and suggested male cranial features were more "developed" than female cranial features and there were pelvic structure differences between the sexes (Hrdlička, 1919). In the second volume of the *American Journal of Physical Anthropology*, Hrdlička (1919) provides a detailed summary of sex estimation methods available from 1866 until 1919. The majority of these references are in French or German, with a focus on the skull, and unfortunately nearly all of them lack English translations. Hrdlička (1919) did echo Dureau's earlier sentiment that sex could be accurately estimated in 80% of cases with just the cranium, 90% with the entire skull, and up to 96% of the time with the entirety of the skeleton. Hooton (1935) published *Development and Correlation of Research in Physical Anthropology* at Harvard University using biological anthropology to identify human remains. In 1939, Krogman published *Guide to the Identification of Human Skeletal Material* in the FBI's *Law Enforcement Bulletin*, one of the first compilations of its kind. This work revolutionized how human remains were examined for identification and brought the concept and utility of skeletal biology to a wider investigative audience. In *Up from the Ape* (1946), Earnest Hooton created an appendix that covered his observations between the differences in male and female skeletons and how one could use these to estimate the sex of unknown individuals based upon the skeletal size and weight, as well as the pelvis (Spencer, 1982). Hooton (1946) noted the following skeletal sex differences: overall size differences in males (e.g., cranial size, supraorbital margins, mental eminence) and the frontal sinus is more intricate in males. Despite noting the differences between the sexes, Hooton's work did not include a method per se.

Research in sex estimation

During the early 1900s, methods were being developed primarily by anatomists and biological anthropologists (e.g., Derry, 1909; Martin, 1928; Schultz, 1930). After the 1940s, skeletal biologists (in anatomy, biological anthropology, forensic anthropology, and bioarchaeology) began utilizing the two major skeletal collections in the United States—the Hamann-Todd Human Osteological and Robert J. Terry Anatomical collections—to advance research on identification and sex estimation. However, a need quickly arose for identifying US service members that were skeletonized or badly decomposed. Given that most of the unidentified were males, research at this time focused more on age parameters and stature estimation methods, yet there were some notable additions to the sex estimation literature at this time. In 1953, Hanna and Washburn used ilium, ischium, and pubis measurements and the sciatic notch angle to estimate sex. Starting in 1957, B.J. Boucher began creating methods to estimate the sex of juveniles following methods similar to adult sex estimation. These used metrics of the ilium and morphology of the shape of the greater sciatic notch. Focus at this time was primarily on qualitative features of the pelvis and metrics of the skull. Since the mid-1900s (1950 onward), a multitude of sex estimation methods have been developed, many of which are covered throughout the remaining chapters of this book and which have been well documented elsewhere. See Moore (2013) for a summary of the last 60 years of research in sex estimation and a detailed summary of methods available for each bone.

Conclusion

Before we could ever begin to develop skeletal methods for differentiating the sexes, we first needed to understand human internal anatomy and human osteology from a scientific perspective. What is clear from this historical review is that many of the differences noted early in the literature (e.g., color, weight, number of bones) were inaccurate and not scientifically supported. Forensic applications were at the forefront for moving beyond just describing the sex differences to actually creating usable methods to estimate sex. However, this research was often secondary to the study of ancestry, or delayed due to a focus on other biological profile parameters (e.g., age and stature in the 1900s). The creation of skeletal collections facilitated the development of actual methods for sex estimation, many of which have since been modified but are still employed today. After the 1950s, we see a rapid increase in the number of qualitative and quantitative methods available to estimate sex from the skeleton, many of which are covered throughout this work. Chapter 2 of this volume details the current state of sex estimation methods employed and current practitioner preferences.

References

Ackermann, J. F. (1788). Über die Kreuzung der Sehnerven. In *Blumbach's medical bibliography, 1788.* III. 307. 706.
Adams, F. (Ed.), (1886). *The genuine works of Hippocrates*. New York, NY: William Wood & Company.
Albinus, B. (1749). *Table of the skeleton and muscles of the human body*. London: H. Woodfall.
Bauhin, G., & de Bry, T. (1605). *Theatrum anatomicum*. Frankfort: Matthai Beckeri.

Boucher, B. J. (1957). Sex differences in the foetal pelvis. *American Journal of Physical Anthropology*, *15*, 591–600.

Broca, P. (1875). Instructions craniologiques et craniometriques. In *Bulletins et mémoires de la Société d'anthropologie de Paris*. Paris: Dion.

Connell, S. (2000). Aristotle and Galen on sex difference and reproduction: A new approach to an ancient rivalry. *Studies in History and Philosophy of Science*, *31*(3), 405–427.

Cowper, W. (1737). *The anatomy of humane bodies* (2nd ed.). London: University of Utrecht.

Derry, D. E. (1909). Note on the innominate bone as a factor in the determination of sex. *Journal of Anatomy and Physiology*, *43*(3), 266–276.

Dorsey, G. A. (1897). A sexual study of the size of the articular surfaces of the long bones in aboriginal American skeletons. *Boston Medical and Surgical Journal*, *137*, 80–82.

Dwight, T. (1878). *Identification of the human skeleton: A medico-legal study*. Boston: David Clapp & Son.

Dwight, T. (1894). Range and significance of variation in the human skeleton. *Boston Medical and Surgical Journal*, *131*, 97–101.

Galen, C. G. (AD 1800). De ossibus ad tirones. In C.G, Kühn (Ed.), Opera Omnia. Leipzig: C. Knobloch.

Gardiner, J. R., & Swain, A. (2015). Sex determination. In T. M. Plant & A. J. Zeleznik (Eds.), *Knobil & Neill's physiology of reproduction* (pp. 267–292). New York: Elsevier.

Giles, E., & Elliot, O. (1963). Sex determination by discriminant function analysis of crania. *American Journal of Physical Anthropology*, *21*(1), 53–68.

Hanna, R. E., & Washburn, S. L. (1953). The determination of the sex of skeletons, as illustrated by a study of Eskimo pelvis. *Human Biology*, *25*(1), 21–37.

Harrington, T. F. (1911). An appreciation of Dr. Dwight. *The Sacred Heart Review*, *47*(2), 3–5.

Hooton, E. (1935). Development and correlation of research in physical anthropology at Harvard University. *Proceedings of the American Philosophical Society*, *LXXV*, 499–516.

Hooton, E. A. (1946). *Up from the ape*. New York, NY: The Macmillan Company.

Hrdlička, A. (1919). Anthropometry. *American Journal of Physical Anthropology*, *2*, 43–46.

Krogman, W. M. (1939). *Guide to the identification of human skeletal material. United States. Federal Bureau of Investigation: Law enforcement bulletin.*

Lewis, J. E., DeGusta, D., Meyer, M. R., Monge, J. M., Mann, A. E., et al. (2011). Correction: The mismeasure of science: Stephen Jay Gould versus Samuel George Morton on skulls and bias. *PLoS Biology*, *9*(7). https://doi.org/10.1371/annotation/138c7c99-249f-432c-a4f7-2993b7b87c0a.

Lowell, J. M., Plaisted, H. M., & Maine Supreme Judicial Court. (1875). *Report of the trial of James M. Lowell, indicted for the murder of his wife, Mary Elizabeth Lowell: Before the Supreme judicial court of Maine, for Androscoggin County*. Portland, ME: Dresser, McLellan & Co.

Martin, R. (1928). *Lehrbuch der Anthropologie: In systematischer darstellung mit besonderer berüksichtigung der anthropolgischen methoden für studierende ästze und forschungsreisende* (2nd ed.). Jena: Fisher.

Moore, M. K. (2013). Sex estimation and assessment. In L. DiGangi & M. Moore (Eds.), *Research methods in skeletal biology* (pp. 91–116). New York: Academic Press.

Moreau, E. T. (1750). *Quaestio medica: An praeter genetalia sexus inter se discrepent?* Paris: Quillau.

Persaud, T. V. N. (2014). *A history of human anatomy*. Springfield, IL: Charles C. Thomas.

Portland Transcript. (1874). *Maine matters: Androscoggin County*. [February 14].

Schiebinger, L. (1986). Skeletons in the closet: The first illustrations of the female skeleton in eighteenth-century anatomy. *Representations*, *14*, 42–82.

Schultz, A. H. (1930). The skeleton of the trunk and limbs of higher primates. *Human Biology*, *2*, 303–438.

Singer, C. (1956). *Galen on anatomical procedures: de Anatomicis administrationibus*. London, England: Oxford University Press.

Snow, C. (1982). Forensic anthropology. *Annual Review of Anthropology*, *11*, 97–131.

Spencer, F. (Ed.), (1982). *A history of American physical anthropology 1930–1980*. New York, NY: Academic Press.

Stévant, I., Papaioannou, M., & Nef, S. (2018). A brief history of sex determination. *Molecular and Cellular Endocrinology*, *468*, 3–10.

Stewart, T. D. (1979). *Essentials of forensic anthropology: Especially as developed in the United States*. Springfield, IL: Charles C. Thomas.

Tz'u, S. (1981). *The washing away of wrongs* [B. E. McKnight, Trans.] Center for Chinese Studies: The University of Michigan [Original work published 1247].

Ubelaker, D. (2018). A history of forensic anthropology. *American Journal of Physical Anthropology*, *165*, 915–923.

Ubelaker, D. H., & De La Paz, J. S. (2012). Skeletal indicators of pregnancy and parturition: A historical review. *Journal of Forensic Sciences*, *57*(4), 866–872.

Vesalius, A. (1543). *De humani corporis fabrica libri septem*. Basileae: Anno salutis reparatae.

von Soemmerring, S. T. (1796). *Tabula sceleti feminini juncta descriptione*. Traiecti ad Moenum: Apud Varrentrapp et Wenner. MDCCLXXXXVI.

Woo, E. U., Jung, H., & Tansatit, T. (2018). Cranial index in modern people of Thai ancestry. *Anatomy and Cell Biology*, *51*(1), 25–30.

Zysk, K. (1986). The evolution of anatomical knowledge in ancient India, with special reference to cross-cultural influences. *Journal of the American Oriental Society*, *106*(4), 687–705.

CHAPTER 2

Practitioner preferences for sex estimation from human skeletal remains

Alexandra R. Klales
Forensic Anthropology Program, Department of Sociology and Anthropology, Washburn University, Topeka, KS, United States

Introduction

Methods used to estimate sex are continuously being updated to account for population differences (see Chapter 17 of this volume), secular changes (see Chapter 18 of this volume), and required statistical rigor (see Chapter 13 of this volume); however, Moore (2013, p.112) notes that "research methods in sex estimation … have changed a great deal in the last 50 years, yet many of the observations are still the same." The methods chosen to complete a skeletal estimation of sex depend on the bones available for analyses, the condition of those bones (e.g., alterations from trauma or taphonomy), and the general age and ancestry of the individual. In some forensic cases, it may not be necessary to estimate skeletal sex because of the presence of soft tissue indicators including external genitalia, internal genitalia, and secondary sex characteristics. In other cases, DNA analyses of sex may be conducted; however, the Scientific Working Group for Forensic Anthropology's document *Sex Assessment* (2010) recommends completing a skeletal estimation of sex even if DNA analyses are completed or are going to be completed. As new methods and validations emerge, we know that the pelvis (innominates and sacrum), specifically the pubis, remains the best skeletal indicator of sex, followed by postcranial metrics, and then morphology and metrics of the skull (Klales, Ousley, & Vollner, 2012; Spradley & Jantz, 2011). However, in some instances, these skeletal regions are absent or too badly damaged for analyses, especially in bioarchaeological contexts or in forensic cases that involve extensive burning. In these instances, the analyst must turn to other skeletal elements, and for this reason, nearly every bone of the human body has been studied for its utility in sex estimation.

Historically, qualitative assessments dominated biological anthropology, especially for sex and ancestry; however, there has been a perceived shift in the past several decades toward the development and greater use of quantitative approaches (Dirkmaat, Cabo, Ousley, & Symes, 2008; Moore, 2013). Attempts have been made to standardize the various methods used for biological profile estimation and the data collection process, including Buikstra and Ubelaker's (1994) work *Standards for Data Collection from Human Remains* and more recently with the Osteoware Standardize Skeletal Documentation

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00002-X

Software (Wilczak & Dudar, 2011). Further attempts to create standardized protocols within the field have also since been proposed and facilitated with the creation of the Scientific Working Group for Forensic Anthropology (2010), the anthropology subcommittee of the Organization of Scientific Area Committees in 2014 under the National Institute of Standards and Technology, and the Anthropology Consensus Body of the American Academy of Forensic Sciences Standards Board in 2018. The extent to which standardization has been actually implemented for sex estimation by the biological anthropology community as a whole is currently unknown.

As Garvin (2012, p. 245) notes, "preferred sex [estimation] methods will vary according to the anthropologist's personal preferences and experience." The way in which the results are reported also likely varies considerably by practitioner. In many forensic cases and bioarchaeological reports, both qualitative and quantitative methods are employed for sex estimation to generate the biological profile. The lack of consistency within biological anthropology for the estimation of biological profile parameters is problematic and raises questions of methodological protocol for sex estimation and consistency, especially within forensic anthropology. Further complicating the lack of consistency is the fact that nearly all other biological profile parameters can depend on correct sex estimation.

The goal of this research was to investigate the methodological choices made by biological anthropologists for sex estimation. Understanding the degree of variability, method preference, and modes of reporting is the first step toward standardization within the field. Garvin and Passalacqua (2012, p. 427) documented current practices for adult age estimation and suggest that reporting results of current practice will "raise awareness of our practices as a unified discipline and promote discussion on future improvements and standardization." In an attempt to remain consistent with previous research assessing current practices for adult age estimation, a format similar to that used by Garvin and Passalacqua (2012) was utilized for both data collection and reporting the results, and is described in more detail below.

Materials and methods

An electronic questionnaire was created using the online surveying program Kwiksurvey. The online questionnaire consisted of 32 questions concerning the participant's education, background, and their preferences and practices for sex estimation (Klales, 2013). The survey took approximately 15 min to complete. Like Garvin and Passalacqua's (2012) age estimation survey, a multiple-choice format was utilized when possible; and for the methodological questions, a rank system was used and included an area for additional written comments. Similarly, it was not possible to include all skeletal areas and methods in the survey, so only those that were perceived by the author to be the most popular at the time were included and participants were able to include additional areas and methods in the comments section. Participations were able to opt to answer all of the questions listed or only specific ones within the survey once beginning the questionnaire; therefore, not all questions were answered by each participant.

Participants were recruited via email communication through a bulk list serve distributed based on memberships in professional organizations that include anthropologists, or through announcement of the research on these organizations' websites (e.g., American Academy of Forensic Sciences, American Association of Physical Anthropologists, Canadian Association for Physical Anthropology, American Anthropological Association, European Anthropology Association, etc.). Responses were received from 154 individuals. Participation in this research was completely voluntary, and participants were not compensated. Respondents agreed to be willing participants in the current research, as well as agreed to allow the answers they provided to be used for papers, presentations, and publications. Furthermore, participation in the survey was anonymous, and no identification information, such as name, IP address, or affiliations, was collected. This research and included questionnaire was approved by the Joint-Faculty Research Ethics Board at the University of Manitoba prior to being distributed.

Results

Education

Of the 154 respondents, 53.9% had obtained a doctorate, while 37.6% had obtained a master's degree (MA/MS) specifically in anthropology. A small portion of participants only had a BA/BS (7.1%) or were currently enrolled as an undergraduate (1.3%). An additional 29.2% ($n = 45$) of those respondents also held degrees in fields outside of anthropology, including archaeology, biology, chemistry, psychology, zoology, criminal justice, forensic sciences, health sciences, history, physiology, and classics. Most respondents (81.9%) received their highest degree after 1989 (Fig. 1). Notice that a large increase

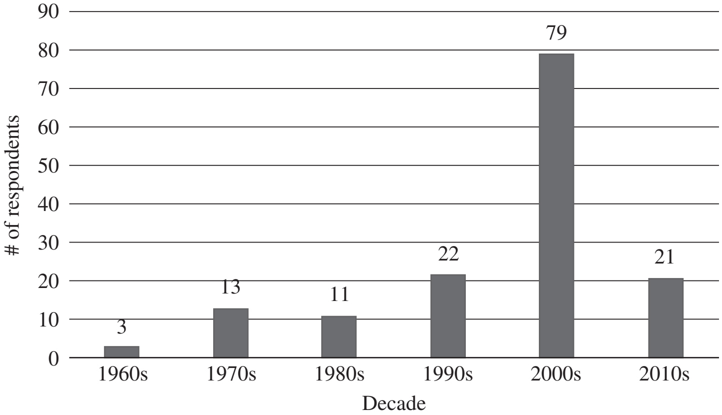


Fig. 1 Year highest degrees in anthropology were conferred.

in degrees awarded occurred in the 2000s, which corresponds to the release of popular crime shows like CSI (first aired 2000) and Bones (first aired 2005).

Current status and affiliations

Participants were asked to define their primary position or current employment and region of permanent operation. Over a third (37.3%) were currently employed in academic settings, as faculty, adjunct faculty, or lecturers. The second largest sector (28.8%) of participants was graduate students, followed by those who identified as other professionals (22.9%). Current professional positions include forensic anthropology consultants, government forensic anthropologists, laboratory directors, medical examiner/coroner employees, museum curators/researchers, medical doctors, and an academic administrator. The remaining participants identified as retirees, independent researches, undergraduate students, and individuals completing a postdoctoral position (Table 1). The majority of individuals listed the United States (59.3%) as the primary location of permanent employment, followed by Canada (20.0%) and Europe (13.3%) (Table 2).

Respondents were next asked to list organization memberships, identify their area of expertise, and to self-identify as specialists. Respondents were members of multiple anthropological organizations, both in the United States and internationally (Table 3), and many individuals were members of multiple professional organizations. When asked "which subfield(s) of anthropology best describes your main field(s) of expertise (check all

Table 1 Current professional positions of participants.

Position	*n*	%
Academic/faculty	57	37.3
Graduate student	44	28.8
Other professional	35	22.9
Independent researcher	8	5.2
Retired	5	3.3
Postdoc	2	1.3
Undergraduate student	2	1.3

Table 2 Current primary locations of participants.

Region	*n*	%
United States	89	59.3
Canada	30	20.0
Europe	20	13.3
Australia	7	4.7
Africa	3	2.0
S. America	1	0.7

Table 3 Organization affiliations of participants.

Organization	*n*
American Association of Physical Anthropologists	92
American Academy of Forensic Sciences	77
Canadian Association for Physical Anthropology	36
Australasian Society for Human Biology	10
Paleopathology Association	7
American Anthropological Association	6
European Anthropological Association	5
British Association for Bioarchaeology and Osteoarchaeology	5
Society of Forensic Anthropologists	3

Table 4 Self-identified areas of expertise of participants.

Main area(s) of expertise	*n*	%
Physical/biological only	115	76.7
Biological/archaeology	26	17.3
Biological/medical	4	2.7
Biological/archaeology/cultural	2	1.3
Biological/cultural/medical	1	0.7
Archaeology only	1	0.7
Cultural/linguistics	1	0.7

that apply)," the greatest number of responders (76.7%) identified solely as physical or biological anthropologists, while 17.3% described their main fields of expertise as both physical/biological anthropology and archaeology (Table 4). Over two-thirds of respondents identified as bioarchaeologists (65.6%) and/or as forensic anthropologists (60.9%).

Experience

Most responders (82.1%) began conducting osteological research, including sex estimation, between 1980 and 2009 (Fig. 2). Participants were then asked to describe the number of cases in which they had conducted osteological analysis, including estimation of sex, in both bioarchaeological and forensic contexts. Within bioarchaeology, the highest number of responders (26.2%) reported completing analyses for 501–1000 cases (Fig. 3). Within forensic anthropology, the level of experience with cases varied. The majority of participants (46%) had experience with less than 100 forensic cases, while nearly a third (28.7%) had no experience with active forensic cases (Fig. 4).

Relationships between degree attained, years of experience, and number of both types of cases were explored using Spearman's rank correlations. An increase in the number of PhDs is noted in the 1980s and 1990s (Fig. 5). Not surprisingly, respondents with higher degrees had more bioarchaeological ($r = 0.431$, $P < .01$) and forensic case

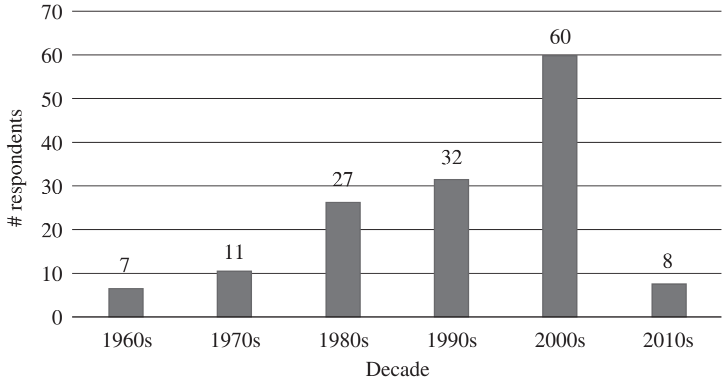


Fig. 2 Decade participants began conducting sex estimation.

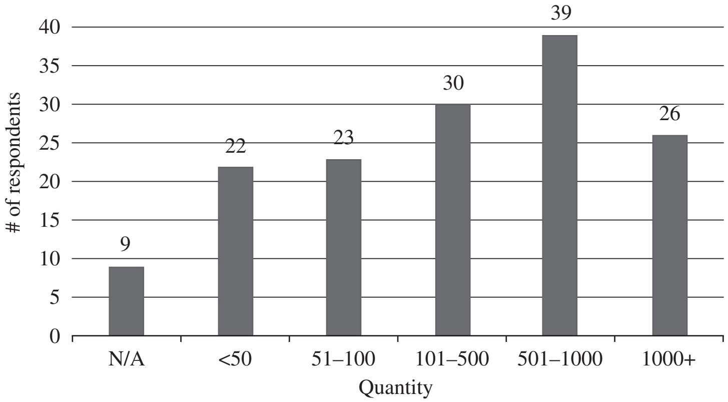


Fig. 3 Number of bioarchaeology cases.

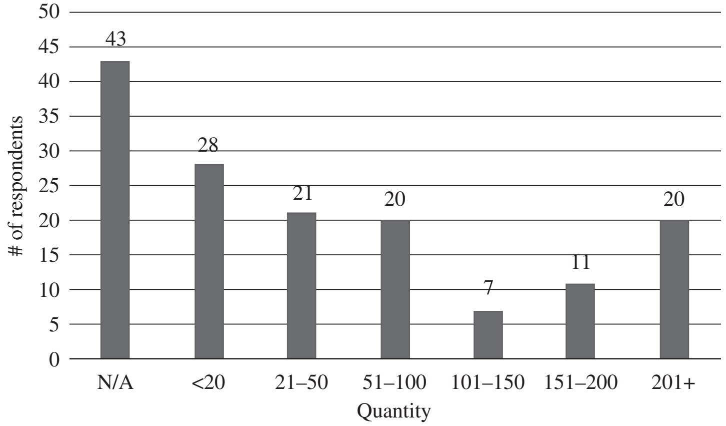


Fig. 4 Number of active forensic cases.

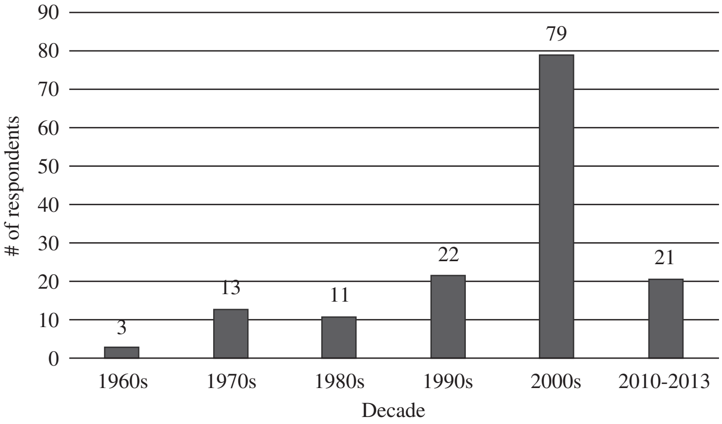


Fig. 5 Number of PhDs attained by decade.

experience ($r = 0.357$, $P < .01$). Total years of experience, based on when they began practicing, correlated to a higher number of bioarchaeology cases ($r = 0.455$, $P < .01$) and forensic anthropology cases ($r = 0.393$, $P < .01$).

Sex estimation preferences

Participants were first asked about their methodological preferences for sex estimation between metric/quantitative and morphological/qualitative. Two-thirds of participants used both method types to estimate sex with a complete skeleton (62.6%). When not using both methods, qualitative methods were preferred over 2 to 1 (25.9% vs 11.5%). Given the push for standardization in the field and the perceived greater objectivity of quantitative methods (Dirkmaat et al., 2008), this was somewhat surprising. However, ease of use and tradition may explain why morphological methods continue to be utilized.

Next, participants were asked to rank the skull, pelvis, long bones, and the hands and feet, based on their preference and perceived reliability of those areas for sex estimation. Skeletal regions were ranked on a scale from 1 to 4, with 1 being the most useful or valid and 4 being the least useful or valid according to the respondent. Participants were also given the option of including an "other" skeletal region in which they could describe another area of preference. In instances when other was selected, participants ranked skeletal regions on a scale from 1 to 5. The lower the average rank score, the higher the use and utility according to the respondents.

The pelvis overwhelmingly ranked first in preference (average rank 1.2), followed by the skull (average rank 2.2), and long bones (average rank of 2.9) (Table 5). Average rank scores for skeletal region preference were also calculated based on the decade the respondent began conducting osteological research, including sex estimation. Regardless of the

Table 5 Preferred skeletal regions for sex estimation.

Region	First	Second	Third	Fourth	Fifth	Don't use	Average
Pelvis	123	10	0	0	4	0	1.2
Skull	10	94	29	3	0	1	2.2
Long bones	0	27	98	8	0	4	2.9
Hands/feet	3	1	3	80	41	9	4.2
Other	2	4	1	30	47	53	4.4

level of experience, based on years of practice, the pelvis was always the most preferred, followed by the skull. When the pelvis was not ranked first, it was ranked second after the skull, which is surprising given the abundance of literature suggesting that the pelvis is the most sexually dimorphic region of the skeleton. Research by Spradley and Jantz (2011) published just prior to the administration of this survey suggests that long bone metrics actually outperform the skull in correct sex classification. In most cases when "other regions" of the skeleton were utilized for sex estimation ($n=84$), the rank score was typically fourth ($n=30$) or fifth ($n=47$). For 18.4% ($n=16$) of participants, "other" skeletal areas were preferred over the hands and feet, while 31.0% ($n=27$) preferred to use other skeletal areas after the hands and feet. Popular answers in the "other" category included (in the order of most listed to least listed): clavicle/scapula, vertebrae, sternum/teeth, hyoid/patella. Two participants included overall size and robusticity of the skeleton as a last preference, and one individual included cartilage ossification.

Finally, participants were asked to rank the quantitative and qualitative methods, sources, or traits they employed for each skeletal region based on their preference, with 1 being the preferred method and so forth. It was not possible to make the list of sources exhaustive, so participants were given the option of including an "other" category and were asked to provide the name and year of the source that they preferred. In the case of including the other category, participants were able to include that in their ranking. It should be noted that some of the methods or sources were not mutually exclusive and included compilations of traits from multiple sources. For example, Buikstra and Ubelaker's (1994) qualitative skull traits are the same used by Walker (2008); however, Walker included these traits within a statistical framework for classification. In these cases, it is hoped that participants ranked their preference for the source itself rather than the traits; however, this cannot be discerned from the data. Average rank is based on the total number of participants who use the method. This ranking approach also assumes that all skeletal areas are present for analysis.

Skull

All participants who completed this question ($n=120$), with the exception of one, use morphological traits of the skull for sex estimation. The five traits listed in Buikstra and

Table 6 Preferred methods, traits, or sources for sex estimation of the skull based on morphology.

Source	First	Second	Third	Fourth	Fifth	Sixth	Average rank
Buikstra and Ubelaker (1994)	71	32	10	5	1	0	1.6
Walker (2008)	23	42	20	8	5	2	2.4
Williams and Rogers (2006)	9	10	23	15	23	2	3.5
Krogman and Iscan (1986)	6	11	31	31	12	4	3.5
Acsádi and Nemeskéri (1970)	2	10	13	22	30	5	4
Other	3	4	1	1	0	10	

Ubelaker (1994) (nuchal crest, mental eminence/trigon, supraorbital margin, supraorbital ridge/glabella, mastoid process) are the most preferred and are utilized by all participants who responded to this question (average rank 1.6) (Table 6). 59.7% of participants selected this source as their first choice overall, while 26.9% selected it as their second choice. Walker's (2008) use of those traits within a statistical framework was preferred as the second most utilized source (average rank 2.4) and was the highest selected source for second preferred method (35.3%). A small number (16.0%) of participants included other methods. Two participants cited Bass (2005), while the remaining respondents cited different population-specific methods, such as Larnach and Freedman (1964) or Maat, Mastwijk, and Van der Velde (1997) and De Villiers (1968). One person indicated that they only use the skull if that is all that is available, and another indicated that they used their own reference data and methods.

15.3% of participants do not use or prefer metric methods for estimating sex with the skull ($n = 124$). For those respondents utilizing metric methods, FORDISC 3.0 (Jantz & Ousley, 2005) was their first choice (70.4%) with an average rank score of 1.7 (see Chapter 12 of this volume for more detail on the program). Of that 70.4%, 18.8% use only FORDISC and no other methods of metric sex estimation with the skull. Howells's (1973) discriminant function scores ranked second highest (average rank 2.7), with 4.9% ranking it as their first choice and 33.0% of participants choosing it as the second preferred method. The use of other computer programs, like CRANID (average rank 3.4) and 3D-ID ($n = 2$), ranked low. The following methods were listed by one person each: Spradley and Jantz (2011), Albanese (2013), and Dayal, Spocter, and Bidmos (2008). Surprisingly, two of these references focus more so on the postcrania than the skull; however, because they were listed on the survey in this section they are included here.

Table 7 Preferred methods, traits, or sources for sex estimation of the pelvis based on morphology.

Source	First	Second	Third	Fourth	Fifth	Sixth	Average rank
Phenice (1969)	61	26	16	5	2	2	1.8
Buikstra and Ubelaker (1994)	35	50	21	2	5	2	2.0
Krogman and Iscan (1986)	5	11	29	38	10	1	3.4
Rogers and Saunders (1994)	7	17	19	24	13	7	3.5
Parturition scars	3	8	11	10	34	9	4.2
Other	6	0	2	1	2	6	

Pelvis

All participants who completed this question ($n = 117$), with the exception of one, use morphological traits of the pelvis for sex estimation. The three traits listed by Phenice (1969) (ventral arc, subpubic concavity, medial aspect of the ischio-pubic ramus) are the most preferred (average rank 1.8) and are utilized by all participants who responded to this question (Table 7). 52.1% of participants selected this source as their first choice overall, while 22.2% selected it as their second choice. The integration of these three traits with the greater sciatic notch and preauricular sulcus in Buikstra and Ubelaker's (1994) *Standards* ranked second (average rank 2.0) and were selected as the second most utilized source (29.9%). 14.5% of participants included other methods. Five participants cited Bruzek (2002), while the remaining "others" included Bass (2005), Brothwell (1981), White and Folkens (2005), Walker (2005), Ascádi and Nemeskéri (1970), and overall size, shape, and robusticity.

33.3% of participants do not use or prefer metric methods for estimating sex from the pelvis ($n = 123$). For those respondents utilizing metric methods, FORDISC 3.0 (Jantz & Ousley, 2005) was their first choice (41.5%), with an average rank score of 1.7. Of that 41.5%, 21.6% use only FORDISC and no other methods of metric sex estimation with the pelvis. Geometric morphometric analyses ranked second highest (average rank 2.7), followed by the ischio-pubic index (average rank 2.9). Three participants indicated their preference for metric assessment of the pelvis using the DSP program (Brůžek, Santos, Dutailly, Murail, & Cunha, 2017) (see Chapter 15 of this volume for more detail on the program).

Long bones

Most participants responded (66.6%) that they prefer not to use morphological traits of the postcrania, aside from the pelvis, for sex estimation. Those who do, use overall size and robusticity (average rank 1.5) and Roger's (1999) traits of the distal humerus (average rank 1.8). 9.8% of respondents do not metrically asses postcranial long bones. For those who do, most prefer FORDISC 3.0 (Jantz & Ousley, 2005) as either first choice (59.3%) or second choice (11.6%). Other popular responses were femoral head diameter (average rank 2.2) and humeral head diameter (average rank 3.3). In the other category, Spradley and Jantz (2011) and Murail, Bruzek, Houët, and Cunha (2005) were each included once, as was "I use my own standards based on local skeletal collections."

Reporting

Lastly, participants were asked if they report the results of the methods they ranked in: research publications (yes: 88.7%), archaeological site reports (yes: 71.3%), and forensic case reports (yes: 66.9%). When multiple methods did not agree on sex estimation, 42.0% present results for each method, 29.8% give preference to one skeletal region or method over others, 12.2% decide based on their own personal experience and general impressions, 12.2% take the average of all methods, and lastly 3.8% present each method but present a final estimation based on their professional opinion and experience.

Conclusion

A recent study by Thomas, Parks, and Richard (2016) compared the accuracy rates between sex estimation using traditional biological anthropology approaches (metric and morphological) to known sex from DNA analyses. Total accuracy was 94.7% ($n=360$ cases), with accuracy increasing as more of the skeletal remains were available for analysis. With a complete skeleton, sex estimation accuracy was 97.8% (Thomas et al., 2016). The lowest accuracy was in cases with only the mandible present (60.0%) or long bones (77.8%). Of the 5.2% ($n=19$) incorrectly sexed cases in their analyses, 60% ($n=9$) of these only had a single element available for analysis. Other individual studies have also compared estimated sex to determined sex from DNA in Sweden (Rogers, 1999), Russia (Ovchinnikov, Ovtchinnikova, Druzina, Buzhilova, & Makarov, 1998), Turkey (Matheson & Loy, 2001), and Germany (Hummel, Bramanti, Finke, & Herrmann, 2000) with somewhat similar results. These studies demonstrate the accuracy of currently available sex estimation methods and support the recommendation that the entirety of the skeleton should be used for sex estimation when possible. However, their results cannot be used as a *post facto* explanation of error rates in the field (Christensen, Passalacqua, & Bartelink, 2014; Thomas et al., 2016), and validity and reliability tests must continue.

As with the age study by Garvin and Passalacqua (2012), there is considerable variation in practitioner preferences for sex estimation. The preference for FORDISC 3.0 (Jantz & Ousley, 2005) for quantitative analyses is unsurprising given the pervasiveness of the program and its inclusion of both modern and historic samples. Also, quite unsurprising is the variation in qualitative method preferences, as much of this depends on training and lineage. What *is* surprising—and perhaps quite alarming at least from a forensics perspective—is the number of respondents who reported using their own data, own reference collections, unpublished methods, and/or invalidated and unreliable methods for sex estimation. With this new understanding of preferences from this research, we can take the first steps toward standardization of methodology and move away from individual preferences for skeletal sex estimation.

References

Acsádi, G., & Nemeskéri, J. (1970). *History of human life span and mortality*. Budapest: Akadémiai Kiadó.

Albanese, J. (2013). A method for estimating sex using the clavicle, humerus, radius, and ulna. *Journal of Forensic Sciences*, *58*(6), 1413–1419.

Bass, W. M. (2005). *Human osteology: A laboratory and field manual* (5th ed.). Columbia, MO: Missouri Archaeological Society.

Bruzek, J. (2002). A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology*, *117*, 157–168.

Brůžek, J., Santos, F., Dutailly, B., Murail, P., & Cunha, E. (2017). Validation and reliability of the sex estimation of the human os coxae using freely available DSP2 software for bioarchaeology and forensic anthropology. *American Journal of Physical Anthropology*, *164*(2), 440–449.

Brothwell, R., Dr. (1981). *Digging up bones* (3rd ed.). Ithaca, New York: Cornell University Press.

Buikstra, J. E., & Ubelaker, D. H. (1994). Standards for data collection from human skeletal remains. In *Archaeological survey research seminar series 44, Fayetteville, AR*.

Christensen, A. M., Passalacqua, N. V., & Bartelink, E. J. (2014). *Current methods in forensic anthropology* (1st ed.). United Kingdom: Elsevier.

Dayal, M. R., Spocter, M. A., & Bidmos, M. A. (2008). An assessment of sex using the skull of black South Africans by discriminant function analysis. *HOMO*, *59*(3), 209–221.

De Villiers, H. (1968). Sexual dimorphism of the skull of the South African Bantu speaking Negro. *South African Journal of Science*, *64*, 118–124.

Dirkmaat, D. C., Cabo, L. L., Ousley, S. D., & Symes, S. A. (2008). New perspectives in forensic anthropology. *Yearbook of Physical Anthropology*, *137*, 33–52.

Garvin, H. M. (2012). Adult sex determination: Methods and application. In D. C. Dirkmaat (Ed.), *A companion to forensic anthropology* (pp. 239–247). West Sussex: Wiley-Blackwell.

Garvin, H. M., & Passalacqua, N. V. (2012). Current practices by forensic anthropologists in adult skeletal age estimation. *Journal of Forensic Sciences*, *57*, 427–433.

Howells, W. W. (1973). *Cranial variation in man. Vol. 67*. Cambridge, MA: Papers of the Peabody Museum.

Hummel, S., Bramanti, B., Finke, T., & Herrmann, B. (2000). Evaluation of morphological sex determinations by molecular analyses. *Anthropologischer Anzeiger*, 9–13.

Jantz, R. L., & Ousley, S. D. (2005). *FORDISC 3.0. computer program*. Knoxville, TN: University of Tennessee Knoxville.

Klales, A. R. (2013). Current practices in forensic anthropology for sex estimation in unidentified, adult individuals. In: *Proceedings of the 65th annual meeting of the American Academy of Forensic Sciences; 2013 Feb 18–23*, Washington, DC, Colorado Springs, CO: American Academy of Forensic Sciences, 19(H81), 439–440.

Klales, A. R., Ousley, S. D., & Vollner, J. M. (2012). A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *American Journal of Physical Anthropology*, *149*, 104–114.

Krogman, W. M., & Iscan, M. Y. (1986). *The human skeleton in forensic medicine* (2nd ed.). Springfield: Charles C. Thomas.

Larnach, S. L., & Freedman, L. (1964). Sex determination of Aboriginal crania from Coastal New South Wales, Australia. *Record of the Australian Museum*, *26*(11), 295–308.

Matheson, C. D., & Loy, T. H. (2001). Genetic sex identification of 9400-year-old human skull samples from Çayönü Tepesi, Turkey. *Journal of Archaeological Science*, *28*(6), 569–575.

Maat, G., Mastwijk, R., & Van der Velde, E. (1997). On the reliability of non-metrical morphological sex determination of the skull compared with that of the pelvis in The Low Countries. *International Journal of Osteoarchaeology*, 7(6), 575–580.

Moore, M. K. (2013). Sex estimation and assessment. In L. DiGangi & M. Moore (Eds.), *Research methods in skeletal biology* (pp. 91–116). New York: Academic Press.

Murail, P., Bruzek, J., Houët, F., & Cunha, E. (2005). DSP: A tool for probabilistic sex diagnosis using worldwide variability in hip-bone measurements. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, *17*(3–4), 167–176.

Ovchinnikov, I. V., Ovtchinnikova, O. I., Druzina, E. B., Buzhilova, A. P., & Makarov, N. A. (1998). Molecular genetic sex determination of Medieval human remains from North Russia: Comparison with archaeological and anthropological criteria. *Anthropologischer Anzeiger*, 7–15.

Phenice, T. W. (1969). A newly developed visual method for sexing the *os pubis*. *American Journal of Physical Anthropology*, *30*, 297–302.

Rogers, T. L. (1999). Sex determination of adolescent skeletons using the distal humerus. *American Journal of Physical Anthropology*, *140*, 143–148.

Rogers, T., & Saunders, S. (1994). Accuracy of sex determination using morphological traits of the human pelvis. *Journal of Forensic Sciences*, *39*, 1047–1056.

Scientific Working Group for Forensic Anthropology (SWGANTH). (2010). *Sex assessment*. Washington, DC: NIST.

Spradley, M. K., & Jantz, R. L. (2011). Sex estimation in forensic anthropology: Skull versus postcranial elements. *Journal of Forensic Sciences*, *56*(2), 289–296.

Thomas, R. M., Parks, C. L., & Richard, A. H. (2016). Accuracy rates of sex estimation by forensic anthropologists through comparison with DNA typing results in forensic casework. *Journal of Forensic Sciences*, *61* (5), 1307–1310.

Walker, P. L. (2005). Greater sciatic notch morphology: Sex, age, and population differences. *American Journal of Physical Anthropology*, *127*, 385–391.

White, T. D., & Folekns, P. A. (2005). *Human bone manual. Burlington*. Elsevier: MA.

Walker, P. L. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, *136*, 39–50.

Wilczak, C. A., & Dudar, J. C. (Eds.), (2011). *Vol. I. Osteoware^{TM} Software manual*. Washington, DC Smithsonian Institution.

Williams, B. A., & Rogers, T. (2006). Evaluating the accuracy and precision of cranial morphological traits for sex determination. *Journal of Forensic Sciences*, *51*, 729–735.

CHAPTER 3

Applications of sex estimation in paleoanthropology, bioarchaeology, and forensic anthropology

Jonathan D. Bethard[a], **Caroline VanSickle**[b]
[a]Department of Anthropology, University of South Florida, Tampa, FL, United States
[b]Department of Anatomy, A.T. Still University of Health Sciences, Kirksville, MO, United States

Introduction

Biological anthropologists encounter skeletonized remains of modern humans and our ancestors from time periods spanning thousands of millennia and from almost all geographic regions of the world. Discoveries range from fossilized remains of extinct Pliocene hominins like "Lucy" and "Ardi" (Johanson, White, & Coppens, 1978; White et al., 2009), to more historic mortuary contexts only a few centuries removed from the present day (Blakey, 1998; Mack & Blakey, 2004). In addition, some biological anthropologists routinely encounter skeletal remains from modern forensic contexts and are required to perform analyses relevant to the process of human identification (Krishan et al., 2016). In all of these cases, and countless other examples drawn from biological anthropology, sex is a primary piece of information that researchers seek to tease out from the skeletonized remains on their laboratory table. For a discussion of how sex can be estimated from a human skeleton, see related chapters in this volume (Chapters 6–12). In this chapter, we discuss why paleoanthropologists, bioarchaeologists, and forensic anthropologists are interested in estimating sex from skeletal remains and the limitations they each face in doing so.

Skeletal sex estimation in paleoanthropology

Paleoanthropologists estimate sex from the skeletal remains of fossil hominins to address evolutionary-focused questions relating to sexual dimorphism, including how hominin social systems changed over time (Darwin, 1888; Lovejoy, 1981; Plavcan, 2011) or how the pelvis evolved adaptations for parturition as brain size increased in the hominin lineage (Häusler & Schmid, 1995; Rosenberg & Trevathan, 1995; Tague & Lovejoy, 1986; Tague & Lovejoy, 1998). Addressing these questions is an important step in reconstructing the past lifeways of our hominin ancestors. However, sex estimation techniques

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00003-1

developed for use on modern humans are not guaranteed to be applicable to fossil species. Further limitations for paleoanthropologists include distinguishing sex differences from species differences and working with a small and fragmentary sample.

In light of these limitations, paleoanthropologists often must be more general about their sex estimations of fossil remains—instead of having multiple skeletal features on which to base an estimation, they may only have one or two, yielding lower certainty from their results. Similarly, species differences may mean that traits that are sexually dimorphic in humans may not have been sexually dimorphic in past species. Indeed, many paleoanthropology analyses estimate sex based on measures of body size, which requires the assumption that the species being studied was sexually dimorphic and that enough evidence exists to determine the range of variation to distinguish two morphologies. These limitations mean that there are many cases in paleoanthropology where the sex of a fossil individual is debated. One example that highlights these issues comes from *Homo erectus*, where the sex estimation debate intersects with debate over taxonomy.

A broad, lumper's viewpoint of *H. erectus* includes fossils stretching from Tanzania to Indonesia, spanning a period from approximately 1.8 million to 100,000 years ago, and represented by multiple individuals from many different archaeological sites (Antón, 2003). Being a species that lived across a large geographic area for such a long period of time, there is variation between sites (which some interpret as differences between closely related species of *H. erectus*, *H. ergaster*, and *H. georgicus*; here we will refer to these all as *H. erectus*). Postcranial remains of *H. erectus* are relatively rare because the species is best identified by craniodental remains, and the few postcranial remains that exist from the same timespan as *H. erectus* are not associated with craniodental material. The result is that their taxonomy is heavily debated.

This is unfortunate since methods of estimating sex from craniodental remains in modern humans are not directly applicable to *H. erectus*, a species that has a more robust and overall smaller cranium than modern humans. Sex estimates based on *H. erectus* craniodental remains are generally limited to size differences, which vary site-to-site (Antón, 2003). In modern humans, pelvic features are considered a reliable method of estimating sex (see Chapter 6 of this volume) due to an evolutionary history of adaptations for parturition in light of encephalization. The fossil pelvic remains attributed to *H. erectus* (including some that are taxonomically debatable) have modern human-like ilia and ischia (in many of these fossils the pubis and sacrum are broken or absent, limiting the comparisons possible for those skeletal elements), suggesting that they might be subject to similar sexual dimorphism as modern human pelves. Indeed, there are *H. erectus* fossils that have a wide greater sciatic notch (e.g., OH 28 and BSN49/P27—a morphology associated with female pelves in modern humans) and others with a narrow greater sciatic notch (e.g., KNM-ER 3228 and KNM-WT 15000—a morphology associated with male pelves in modern humans). These morphological differences suggest that greater sciatic notch width may have been sexually dimorphic in this fossil species. Yet, this analysis is complicated by the numerous taxonomic questions that exist for this postcranial sample: OH 28, BSN49/P27, and KNM-ER 3228 are all pelvic remains that do not have clearly associated craniodental remains, which means that they could represent a different species, such as *H. habilis* (no clear evidence exists of what this species' pelvis looked like) or even a species of the genus *Paranthropus* (who co-existed with *H. erectus* and *H. habilis*, but whose pelvis is also not well understood due to lack of evidence; note that some authors refer to this genus as robust *Australopithecus*).

The OH 28 fossil hip bone from Olduvai Gorge in Tanzania is a preserved acetabulum with parts of the ilium and ischium. It is noteworthy for having a large, well-preserved acetabulum, a wide greater sciatic notch, and an incredibly robust iliac pillar (acetabulocristal pillar) that terminates at a broken edge of the iliac blade. Based on this individual's apparent body size (from the acetabulum diameter) and overall robusticity (based on the iliac pillar, which is truly much larger and better defined than is typical for modern human ilia), this individual appears to be male (McHenry, 1991; Simpson et al., 2008; Simpson, Quade, Levin, & Semaw, 2014). Yet, based solely on the wide, symmetrical greater sciatic notch, some consider this individual to be female (Ruff, 2010; Ruff & Walker, 1993). On its own, the sex of this fossil individual seems like an unresolvable debate, but it becomes a more significant issue when we consider the 1.4–0.9 Ma fossil pelvis BSN49/P27 from Gona, Ethiopia. This fossil preserves nearly a complete pelvis (parts of both hip bones, including parts of the pubis, and a partial sacrum). Like OH 28, this pelvis has a wide, symmetrical greater sciatic notch; unlike OH 28, it is a very small individual (based on acetabulum diameter and the articular surfaces of the sacrum) (Simpson et al., 2008, 2014). The BSN49/P27 fossil was found in a location and dated to a time when only two species are known: *H. erectus* and *P. boisei*. Based on its locomotor adaptations for bipedalism, which some think *Paranthropus* would have lacked, the discoverers of this fossil pelvis attributed it to *H. erectus* despite lacking corroborating craniodental evidence (Simpson et al., 2008). But this creates a problem for the *H. erectus* pelvic sample, where sorting by body size means OH 28 is male, and sorting by greater sciatic notch morphology means OH 28 is female and much larger than BSN49/P27. The discoverers of BSN49/P27 subscribe to the former idea (Simpson et al., 2008, 2014), while Ruff (2010) believes that because OH 28 is female, and BSN49/P27 is so very small, it cannot possibly be *H. erectus* since it is highly unlikely that any hominin species would have so much intrasex variation. Ruff (2010) suggests that BSN49/P27 may be *Paranthropus* but is certainly not *H. erectus*. Thus, the taxonomy of a well-preserved hominin pelvis hinges on what sex the less-complete OH 28 is.

Even when paleoanthropologists are able to assure that the fossils in question are the same species, and that they belong to a species that was sexually dimorphic in a way that can be assessed from its skeletal remains, paleoanthropologists, like many biologists, are limited by the binary sex framework (see Chapter 4 of this volume). If fossil sex estimations are based on larger individuals being male and smaller individuals being female (or in the case of OH 28, wide greater sciatic notches being female and narrow ones being male), there are only two possible options for sex. Intersex individuals will never be discovered in the fossil record under this system. It may not be possible to identify intersex individuals from skeletal remains; however, when the scientists who write the evolutionary narrative of our species divide all fossils into male or female, it erases the possibility of even asking how other sexes may have contributed to our evolutionary history. When some parts of the fossil record are so small that a new discovery can call into question the interpretation of the sample that came before, as was the case when BSN49/P27 called into question the claim that OH 28 was female, it is important to remember that these are estimations of sex that are designed to fit into a binary framework. Therefore, caution is called for when paleoanthropologists discuss the implications of a fossil being a particular sex, as sex is particularly difficult to estimate when the range of variation for a species is unknown.

Skeletal sex estimation in bioarchaeology

Though the systematic study of human skeletal remains from archaeological sites is a relatively recent addition to academic scholarship, interest in this class of archaeological mortuary material has persisted for well over two centuries (Buikstra & Beck, 2006; Martin, Harrod, & Pérez, 2013). For example, in the United States, numerous 19th- and early-20th-century contributions explicitly examined skeletal assemblages to investigate the origins of Native Americans (Beck, 2006), and the human skeleton has been used since the 18th century to answer questions related to variation found among *Homo sapiens* (Blumenbach, 1775). More recently, bioarchaeologists have offered novel insights into the history of the human condition through their diverse analyses, particularly over the first decades of the 21st century (Sheridan, 2017; Stojanowski & Duncan, 2015). Intense academic interest in archaeological communities was popularized in the 20th century, and the specialty of bioarchaeology was formally defined by Buikstra, 1977. In the same volume, Christopher Peebles penned the now often-cited observation: "a human burial contains more anthropological information per cubic meter of deposit than any other type of archaeological feature" (Peebles, 1977, p. 124). Since these influential comments in the late 1970s, bioarchaeologists have produced scholarship on a variety of specialties, and the discipline has emerged as a central field within broader anthropological discourse (Agarwal & Glencross, 2011; Baadsgaard, Boutin, & Buikstra, 2012; Buikstra & Beck, 2006; Knudson & Stojanowski, 2009; Knüsel & Smith, 2013; Larsen, 2015).

Regardless of a specific regional or temporal focus, bioarchaeology couples osteological information, such as sex, with contextual information derived from archaeological excavations. Sex estimates are the first of several biological parameters that allow researchers to better understand the paleodemography and sex-specific activities of past populations (Agarwal & Wesp, 2017; Hollimon, 2011; Kelly & Ardren, 2016). In some instances, incorrect sex estimates have fundamentally altered the overall interpretation of an archaeological site's function. A now classic example of this misinterpretation involves the monumental site of Machu Picchu in the Andean region of Peru. Approximately 5 years after explorer Hiram Bingham brought worldwide attention to Machu Picchu, George F. Eaton published a monograph describing the skeletal remains recovered from the site (Eaton, 1916). In the 1916 monograph, Eaton concluded that most of the Machu Picchu sample was female (109 females and 26 males). Eaton's sex estimates were based on skeletal element size and robusticity and were unfortunately not described in detail in the original 1916 monograph. Verano (2003) notes that Eaton's lack of familiarity with Andean skeletal remains may have led to the published sex bias. This skewed sex distribution led Bingham to assert that Machu Picchu was an *aqllawasi*, or specialized type of Inka community comprising women, and that the majority of the interments represented "Virgins of the Sun." Verano (2003) re-analyzed the Machu Picchu skeletal assemblage to re-assess several of Eaton's findings, most notably the demographic composition of the sample. Though Eaton's interpretation was anecdotally questioned by numerous scholars for decades, Verano's (2003) publication was the first to definitively argue that the sex distribution of burials was relatively balanced and that the individuals not exclusively women who comprised one part of specialized Inka statecraft. Moreover, Verano's (2003) findings corroborated the work of others (e.g., Burger & Salazar, 2004; Hyslop, 1990; Turner, Kamenov, Kingston, & Armelagos, 2009; Turner, Kingston, & Armelagos, 2010) who suggested that Machu Picchu was a royal estate for the Inka emperor Pachacuti. Ultimately, in this example, sex estimates were integral to the correct interpretation of this monumental archaeological site during the time of the Inka Empire.

Beyond the Machu Picchu example, how else do bioarchaeologists use biological sex to understand the past? We can look to the Black Death epidemic for insight into how bioarchaeologists make inferences about sex-specific mortality. The Black Death occurred in 14th-century Europe, killing approximately 30%–50% of the population; it has been described as one of the most lethal epidemics of human history (DeWitte, 2009, 2010, 2015, 2018; DeWitte & Kowaleski, 2017; DeWitte & Wood, 2008). Bioarchaeologist Sharon DeWitte has extensively analyzed sex data from the East Smithfield Black Death cemetery, as well as several pre- and post-Black Death cemeteries, to understand the effects of sex on both frailty and mortality in medieval populations in England (DeWitte, 2009, 2010, 2018). She concluded that sex did not affect the risk of mortality during the epidemic and corroborated historical sources that stated that the Black Death was an indiscriminate killer. However, DeWitte (2010) coupled sex data with osteological indicators of stress (e.g., periosteal new bone formation, cranial porosities, and linear enamel hypoplasia) and found that previous exposure to physiological stress increased the risk of death for men, but not for women. Finally, DeWitte (2018) concluded that females may have been better shielded from deleterious environmental conditions as children prior to the Black Death and, therefore, experienced better survivorship during the epidemic. DeWitte's longstanding research program on the Black Death underscores the importance of skeletal sex estimates for making inferences about mortality in the past.

Bioarchaeologists are also interested in understanding the ways in which both sex and gender created and maintained social hierarchies in the past. In these instances, bioarchaeologists utilize skeletal sex in combination with artifacts associated with individual burials, along with other types of mortuary data (e.g., burial locations) to cautiously identify and interpret gender categories in the past (Agarwal & Wesp, 2017). An example of this type of study involves Cahokia, a large Mississippian-era city located near St. Louis, Missouri. Cahokia was inhabited between the 8th and 15th centuries AD and has been the subject of longstanding archaeological investigations (Pauketat, 2004). In particular, excavations from Mound 72, one of many earthen burial mounds constructed at Cahokia, have provided insight into status differences related to sex and gender in the Cahokia community. Mound 72 contained 25 distinct burial features, including one with two high-status individuals buried with over 10,000 shell beads arranged in a zoomorphic pattern, and another that was a mass grave containing several decapitated individuals and others who presented projectile point injuries (Fowler, Rose, Vander Leest, & Ahler, 1999). Ambrose, Buikstra, and Krueger (2003) used sex and isotopic data from a subsample of the 272 individuals buried in numerous spatially distinct mortuary features at Mound 72 to make inferences about gender-based social stratification. They coupled data regarding extra-local artifacts found with some individuals, nonspecific indicators of stress, and differences in carbon and nitrogen isotopes to conclude that one group of female individuals was of lower social status and had less access to high-quality dietary protein sources (Ambrose et al., 2003). Ultimately, the case study from Mound 72 reiterates that bioarchaeologists are able to reconstruct the ways in which sex and gender impacted past communities. In sum, bioarchaeologists use skeletal sex estimates to identify demographic differences in the past, test hypotheses about sex differences in mortality, and uncover social stratification that relates to sex and gender. While sex estimates can be a powerful tool for reconstructing the ways in which sex and gender impacted past communities, these interpretations are still based on a sex binary model that does not account for all sex differences.

Skeletal sex estimation in forensic anthropology

In addition to applications in paleoanthropology and bioarchaeology, sex estimation is fundamental to forensic anthropology as well. Forensic anthropologists use sex estimates of decomposed and skeletonized human remains to provide useful information for members of the medicolegal community (Berg, 2013; Bruzek & Pascal, 2006;

Christensen, Passalacqua, & Bartelink, 2014; Garvin, 2012; Klepinger, 2006; Komar & Buikstra, 2008; Rowbotham, 2016) and typically perform sex estimation with an accuracy rate of nearly 95% when more than one skeletal element is available for analysis (Thomas, Parks, & Richard, 2016). Sex estimates help law enforcement, medical examiners, coroners, and medicolegal death investigators narrow down the list of potentially unidentified persons by automatically excluding a large percentage of the population. For example, there are over 12,460 individuals listed in "Unidentified Persons" section of the National Missing and Unidentified Persons System (NamUs) (www.namus.gov) who are listed as either male or female. One example is NamUS case #UP61454 and represents an unidentified female recovered from Detroit, Michigan. Any males reported missing from near this region are automatically excluded from further investigation once the sex of this individual has been estimated as female. Conversely, potential reported missing female individuals from the region may warrant further investigation into this case to see if they are a match. In cases like this NamUS example, sex data is a powerful investigative tool; however, forensic anthropologists know very little about sex estimates in cases of intersex or transgender decedents, and no reference data have ever been produced to address this issue (Buchanan, 2014; Geller, 2009). While forensic anthropologists have not addressed this complex problem in the literature, initiatives such as the formation of the Trans Doe Task Force (http://transdoetaskforce.org/) have started to bring awareness to the issue of unidentified transgender decedents.

Sex estimates have also been critical in demonstrating that human rights abuses have taken place and targeted one sex over the other in various locations around the world (Baraybar & Gasior, 2006; Jantz, Kimmerle, & Baraybar, 2008; Kimmerle, Konigsberg, Jantz, & Baraybar, 2008; Klinkner, 2008; Szleszkowski, Thannhäuser, Szwagrzyk, & Jurek, 2015). For instance, Klinkner (2008) recounted the 1995 Srebrenica massacre in which thousands of men and boys were preferentially targeted by the Bosnian Serb Army and murdered. In the aftermath, the trial, which took place in The Hague as part of the International Criminal Tribunal for the former Yugoslavia (ICTY), affirmed that genocide had occurred. Sex data, partially contributed by forensic anthropologists working for the ICTY, comprised some of the evidence that resulted in this finding (Klinkner, 2008). Sex estimation is a critically important component of forensic anthropology, both in the United States and abroad. Indeed, Bethard and DiGangi (2019) have noted that sex estimation has comprised 14.7% ($n = 112$) of all forensic anthropology contributions in the *Journal of Forensic Sciences* published during the first two decades of the 21st century. Only publications focused on age-at-death estimation have received slightly more attention ($n = 119$) during the same time period. Skeletal sex estimates are critical to forensic anthropological contexts; however, the field of forensic anthropology has much more to accomplish regarding identifying intersex or transgender individuals.

Conclusion

Sex estimation can be a powerful tool for describing variation and adaptations in fossil species, uncovering past social stratifications relating to sex or gender in human societies, or identifying decedents in medicolegal contexts. However, sex estimation can only provide a limited window into the past lives of the individuals under analysis. Will we ever know the role intersex individuals played in hominin evolution? Or whether sex differences in archaeological contexts truly provide an accurate perspective on gender roles in the past? Can forensic anthropologists use their skills to identify transgender decedents? The answer to all of these questions may be no, at least for right now. Yet, it is our role as scientists to recognize the questions that our evidence cannot answer and incorporate that limitation into how we interpret the evidence to address questions we can answer. Ignoring these issues reinforces inaccurate ideas about the nature of and relationship between sex and gender. As anthropologists, we must strive to shift to a paradigm of sex estimation that incorporates all the ways of being human, past and present.

Acknowledgment

The authors would like to thank Elizabeth Miller for her encouragement.

References

Agarwal, S. C., & Glencross, B. A. (2011). *Social bioarchaeology*. Malden, MA: John Wiley & Sons.

Agarwal, S. C., & Wesp, J. K. (2017). *Exploring sex and gender in bioarchaeology*. Albuquerque, NM: University of New Mexico Press.

Ambrose, S. H., Buikstra, J., & Krueger, H. W. (2003). Status and gender differences in diet at Mound 72, Cahokia, revealed by isotopic analysis of bone. *Journal of Anthropological Archaeology*, *22*, 217–226.

Antón, S. C. (2003). Natural history of *Homo erectus*. *Yearbook of Physical Anthropology*, *46*, 126–170.

Baadsgaard, A., Boutin, A. T., & Buikstra, J. E. (2012). *Breathing new life into the evidence of death: Contemporary approaches to bioarchaeology*. Santa Fe, NM: School for Advanced Research Press.

Baraybar, J. P., & Gasior, M. (2006). Forensic anthropology and the most probable cause of death in cases of violations against international humanitarian law: An example from Bosnia and Herzegovina. *Journal of Forensic Sciences*, *51*, 103–108.

Beck, L. A. (2006). Kidder, Hooton, Pecos, and the birth of bioarchaeology. In J. E. Buikstra & L. A. Beck (Eds.), *Bioarchaeology: The contextual analysis of human remains*. Burlington, MA: Academic Press.

Berg, G. E. (2013). Determining the sex of unknown human skeletal remains. In M. A. Tersigni-Tarrant & N. R. Shirley (Eds.), *Forensic anthropology: An introduction*. Boca Raton, FL: CRC Press.

Bethard, J. D., & DiGangi, E. A. (2019). From the laboratory to the witness stand: Research trends and method validation in forensic anthropology. In L. C. Fulginit, A. Galloway, & K. M. Hartnett-Mccann (Eds.), *Forensic anthropology and the United States judicial system*. Wiley.

Blakey, M. L. (1998). The New York African Burial Ground Project: An examination of enslaved lives, a construction of ancestral ties. *Transforming Anthropology*, *7*, 53–58.

Blumenbach, J. F. (1775). *De Generis Humani Varietate Nativa*. Jena: Friedrich Schiller University.

Bruzek, J., & Pascal, M. (2006). Methodology and reliability of sex determination from the skeleton. In A. Schmitt, E. Cunha, & J. Pinheiro (Eds.), *Forensic anthropology and medicine: Complementary sciences from recovery to cause of death*. Totowa, NJ: Humana Press.

Buchanan, S. (2014). *Bone modification in male to female transgender surgeries: Considerations for the forensic anthropologist*. MA: Louisiana State University.

Buikstra, J. E. (1977). Biocultural dimensions of archaeological study: A regional perspective. In M. L. Blakey (Ed.), *Biocultural adaptation in prehistoric America*. Athens, GA: University of Georgia Press.

Buikstra, J. E., & Beck, L. (2006). *Bioarcheology: The contextual study of human remains*. New York: Routledge.

Burger, R. L., & Salazar, L. C. (2004). *Machu Picchu: Unveiling the mystery of the Incas*. Yale University Press.

Christensen, A. M., Passalacqua, N. V., & Bartelink, E. J. (2014). Sex estimation. In *Forensic anthropology: Current methods and practice*. San Diego, CA: Academic Press.

Darwin, C. (1888). *The descent of man and selection in relation to sex*. Murray.

DeWitte, S. N. (2009). The effect of sex on risk of mortality during the Black Death in London, AD 1349–1350. *American Journal of Physical Anthropology*, *139*, 222–234.

DeWitte, S. N. (2010). Sex differentials in frailty in medieval England. *American Journal of Physical Anthropology*, *143*, 285–297.

DeWitte, S. N. (2015). Setting the stage for medieval plague: Pre-black death trends in survival and mortality. *American Journal of Physical Anthropology*, *158*, 441–451.

DeWitte, S. N. (2018). Stress, sex, and plague: Patterns of developmental stress and survival in pre- and post-Black Death London. *American Journal of Human Biology*, *30*, e23073.

DeWitte, S. N., & Kowaleski, M. (2017). Black death bodies. In *Fragments: Interdisciplinary approaches to the study of ancient and medieval pasts* (p. 6). .

DeWitte, S. N., & Wood, J. W. (2008). Selectivity of Black Death mortality with respect to preexisting health. *Proceedings of the National Academy of Sciences*, *105*, 1436–1441.

Eaton, G. F. (1916). *The collection of osteological material from Machu Picchu*. New Haven, CT: Connecticut Academy of Arts and Sciences.

Fowler, M. L., Rose, J. C., Vander Leest, B., & Ahler, S. R. (1999). *Mound 72 area: Dedicated and sacred space in early Cahokia*. Springfield, IL: Illinois State Museum Report of Investigations.

Garvin, H. M. (2012). Adult sex determination: Methods and application. In D. C. Dirkmaat (Ed.), *A companion to forensic anthropology*. Malden, MA: Wiley-Blackwell.

Geller, P. L. (2009). Bodyscapes, biology, and heteronormativity. *American Anthropologist*, *111*, 504–516.

Häusler, M., & Schmid, P. (1995). Comparison of the pelves of Sts 14 and AL288-1: Implications for birth and sexual dimorphism in australopithecines. *Journal of Human Evolution*, *29*, 363–383.

Hollimon, S. E. (2011). Sex and gender in bioarchaeological research. In S. C. Agarwal & B. A. Glencross (Eds.), *Social bioarchaeology*. Malden, MA: Wiley-Blackwell.

Hyslop, J. (1990). *Inka settlement planning*. Austin, TX: University of Texas Press.

Jantz, R. L., Kimmerle, E. H., & Baraybar, J. P. (2008). Sexing and stature estimation criteria for Balkan populations. *Journal of Forensic Sciences*, *53*, 601–605.

Johanson, D. C., White, T. D., & Coppens, Y. (1978). *A new species of the genus* Australopithecus *(primates: Hominidae) from the Pliocene of eastern Africa*. Cleveland, OH: Cleveland Museum of Natural History.

Kelly, S. E., & Ardren, T. (2016). *Gendered labor in specialized economies: Archaeological perspectives on female and male work*. Boulder, CO: University Press of Colorado.

Kimmerle, E. H., Konigsberg, L. W., Jantz, R. L., & Baraybar, J. P. (2008). Analysis of age-at-death estimation through the use of pubic symphyseal data. *Journal of Forensic Sciences*, *53*, 558–568.

Klepinger, L. L. (2006). Assignment of sex. In *Fundamentals of forensic anthropology*. Hoboken, NJ: Wiley-Liss.

Klinkner, M. (2008). Proving genocide? Forensic expertise and the ICTY. *Journal of International Criminal Justice*, *6*, 447–466.

Knudson, K. J., & Stojanowski, C. M. (2009). *Bioarchaeology and identity in the Americas*. Gainesville, FL: University Press of Florida.

Knüsel, C., & Smith, M. (2013). *The Routledge handbook of the bioarchaeology of human conflict*. Routledge.

Komar, D. A., & Buikstra, J. E. (2008). Beginning the identification process developing a biological profile. In *Forensic anthropology: Contemporary theory and practice*. New York: Oxford University Press.

Krishan, K., Chatterjee, P. M., Kanchan, T., Kaur, S., Baryah, N., & Singh, R. (2016). A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. *Forensic Science International*, *261*, 165. e1–165. e8.

Larsen, C. S. (2015). *Bioarchaeology: Interpreting behavior from the human skeleton*. Cambridge: Cambridge University Press.

Lovejoy, C. O. (1981). The origin of man. *Science*, *211*, 341–350.

Mack, M. E., & Blakey, M. L. (2004). The New York African Burial Ground Project: Past biases, current dilemmas, and future research opportunities. *Historical Archaeology*, *38*, 10–17.

Martin, D. L., Harrod, R. P., & Pérez, V. R. (2013). *Bioarchaeology: An integrated approach to working with human remains*. New York: Springer.

McHenry, H. M. (1991). Femoral lengths and stature in Plio-Pleistocene hominids. *American Journal of Physical Anthropology*, *85*, 149–158.

Pauketat, T. R. (2004). *Ancient Cahokia and the Mississippians*. Cambridge: Cambridge University Press.

Peebles, C. S. (1977). Biocultural dimensions of archaeological study: An archaeologist's perspective. In M. L. Blakey (Ed.), *Biocultural adaptation in prehistoric America*. Athens, GA: University of Georgia Press.

Plavcan, J. M. (2011). Understanding dimorphism as a function of changes in male and female traits. *Evolutionary anthropology: Issues, news, and reviews*, *20*, 143–155.

Rosenberg, K., & Trevathan, W. (1995). Bipedalism and human birth: The obstetrical dilemma revisited. *Evolutionary anthropology: Issues, news, and reviews*, *4*, 161–168.

Rowbotham, S. K. (2016). Anthropological estimation of sex. In S. Blau & D. H. Ubelaker (Eds.), *Handbook of forensic anthropology and archaeology*. New York: Routledge.

Ruff, C. (2010). Body size and body shape in early hominins—implications of the Gona pelvis. *Journal of Human Evolution*, *58*, 166–178.

Ruff, C. B., & Walker, A. (1993). Body size and body shape. In A. Walker & R. Leakey (Eds.), *The Nariokotome* Homo erectus *skeleton*. Harvard University Press.

Sheridan, S. G. (2017). Bioarchaeology in the ancient Near East: Challenges and future directions for the southern Levant. *Yearbook of Physical Anthropology*, *162*, 110–152.

Simpson, S. W., Quade, J., Levin, N. E., Butler, R., Dupont-Nivet, G., Everett, M., et al. (2008). A female *Homo erectus* pelvis from Gona, Ethiopia. *Science*, *322*, 1089–1092.

Simpson, S. W., Quade, J., Levin, N. E., & Semaw, S. (2014). The female Homo pelvis from Gona: response to Ruff (2010). *Journal of Human Evolution*, *68*, .

Stojanowski, C. M., & Duncan, W. N. (2015). Engaging bodies in the public imagination: Bioarchaeology as social science, science, and humanities. *American Journal of Human Biology*, *27*, 51–60.

Szleszkowski, Ł., Thannhäuser, A., Szwagrzyk, K., & Jurek, T. (2015). The possibility of establishing causes of death on the basis of the exhumed remains of prisoners executed during the communist regime in Poland: the exhumations at Powązki Military Cemetery in Warsaw. *International Journal of Legal Medicine*, *129*, 801–806.

Tague, R. G., & Lovejoy, C. O. (1986). The obstetric pelvis of AL 288-1 (Lucy). *Journal of Human Evolution*, *15*, 237–255.

Tague, R. G., & Lovejoy, C. O. (1998). AL 288-1—Lucy or Lucifer: Gender confusion in the Pliocene. *Journal of Human Evolution*, *35*, 75–94.

Thomas, R. M., Parks, C. L., & Richard, A. H. (2016). Accuracy rates of sex estimation by forensic anthropologists through comparison with DNA typing results in forensic casework. *Journal of Forensic Sciences*, *61*, 1307–1310.

Turner, B. L., Kamenov, G. D., Kingston, J. D., & Armelagos, G. J. (2009). Insights into immigration and social class at Machu Picchu, Peru based on oxygen, strontium, and lead isotopic analysis. *Journal of Archaeological Science*, *36*, 317–332.

Turner, B. L., Kingston, J. D., & Armelagos, G. J. (2010). Variation in dietary histories among the immigrants of Machu Picchu: Carbon and nitrogen isotope evidence. *Chungara, Revista de Antropologı́a Chilena*, *42*.

Verano, J. W. (2003). Human skeletal remains from Machu Picchu: A reexamination of the Yale Peabody Museum's Collections. In R. L. Burger & L. C. Salazar (Eds.), *The 1912 Yale Peruvian scientific expedition collections from Machu Picchu: Human and animal remains*. New Haven: Yale Peabody Museum.

White, T. D., Asfaw, B., Beyene, Y., Haile-Selassie, Y., Lovejoy, C. O., Suwa, G., et al. (2009). Ardipithecus ramidus and the paleobiology of early hominids. *Science*, *326*, 64–86.

CHAPTER 4

The confusion between biological sex and gender and potential implications of misinterpretations

Evan M. Garofalo[a], **Heather M. Garvin**[b]
[a]Department of Basic Medical Sciences, University of Arizona College of Medicine—Phoenix, Phoenix, AZ, United States
[b]Department of Anatomy, Des Moines University, Des Moines, IA, United States

The terms gender and sex are commonly used interchangeably in the society despite some fundamental differences in their formal definitions—differences that can have major implications when it comes to identifying someone from their skeletal remains. The skeletons that forensic and biological anthropologists work with provide information about a person's *biology*, which may or may not correspond to a person's *identity*, whether referring to how a person self-identifies or how others in the society describe an individual. It is the anthropologist's job to take the scientific biological information and convert that to terms that law enforcement and the general society understand and use to describe an individual, with the hope of identifying someone from a set of skeletal remains. The anthropologist is tasked with forming a bridge between the biological and cultural information, making it vital that practitioners develop a basic understanding of biological variation and in the variety of ways a person identifies and presents themselves socially. As such, the aim of this chapter is to present definitions for common sex and gender terms focusing on the differences between the terms, their use in the field of forensic and biological anthropology, and how confusion between terms can impact analyses. Some confusion is almost guaranteed given our rapidly evolving cultural climate. New terms are developed to better describe individuals' sexual and gender identity, and old terms may take on new meanings. The same goes for the biological aspect of sex; as we learn more about our biology and biological variants, new terms are presented, and old ones modified. Although this chapter does not present a comprehensive list of terms, we hope that it illustrates the complexity in sex terminology, spurs discussion, and reminds practitioners to be cognizant of terminology issues and implications.

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00004-3

Defining sex

If talking about definitions, usually the first place to go to is a dictionary. *Merriam-Webster* defines "sex" as:

(1) either of the two major forms of individuals that occur in many species and that are distinguished respectively as female or male especially on the basis of their reproductive organs and structures,

(2) the sum of the structural, functional, and behavioral characteristics of organisms that are involved in reproduction marked by the union of gametes and that distinguish males and females,

(3) a: sexually motivated phenomena or behavior, b: sexual intercourse, and

(4) genitalia (Sex, n.d.).

From these definitions, we can conclude that "sex" can either refer to an act or a category of individual. We are focusing on the latter in this chapter. The definitions also imply that there are only two categories of sex—female or male—and that classification is based on reproductive organs and gametes. In other words, sex is biological. In humans, specifically, a person is typically born with a set of two X chromosomes (XX) or an X and Y chromosome (XY). Typically, XX individuals develop female reproductive organs, while XY individuals develop male reproductive organs. Note the word "typically" added to the previous sentences. This is because, although it is true in most cases, individuals may have other chromosomal variants (e.g., XXX, XXY, XO), or differences of sexual development resulting in biological outcomes that may not fit neatly into two—male and female—sex categories. This will be discussed in further detail later. Right now, we want to focus on the generalized process of sex differentiation, the fact that a developing fetus cannot control the sex they will be assigned at birth (nor can the parents), and that it is based on biological processes (specifically meiosis, fertilization, and zygote formation). A sex is assigned to each child at birth based on the visual outcome of these processes. The biology of sex is quite complicated and more variable than most people realize, but we will begin with generalizations.

In humans, each cell usually contains 23 *pairs* of chromosomes for a total of 46 chromosomes. One pair of chromosomes are the sex chromosomes. In general, this pair of chromosomes will consist of two X chromosomes (XX) associated with female sex assignment where gametes are called eggs (or ova), and an X and a Y chromosome (XY) associated with male sex assignment where gametes are called sperm. During gamete production, the 23 pairs of chromosomes are split apart during a process known as meiosis. Each egg typically contains a single (unpaired) X chromosome, and each sperm will typically carry either an unpaired X or a Y chromosome. If a sperm carrying an X chromosome succeeds in fertilizing an egg (also carrying an X chromosome), the developing embryo will carry two X chromosomes (XX)—usually associated with female sex differentiation. If a sperm carrying a Y chromosome succeeds, the embryo will have an XY chromosome arrangement—usually associated with male sex differentiation. In conjunction with hormones, sex chromosomes carry genes that regulate the development of reproductive organs. They also either directly or indirectly (via hormone regulation) contribute to the development of other sex traits, including the sex differences that anthropologists observe in the skeleton. For example, features associated with male sex estimates tend to be larger and display more robust skeletal elements, both related to larger size and muscle mass, more prominent cranial features, and pelvic morphologies related to narrower pelvic inlets (see Chapters 6–12 of this volume for detailed skeletal sex differences). These skeletal traits are not binary. In most cases, they are present in individuals assigned male and female sexes, with a degree of overlap in expressions; however, the degree of trait expression or specific morphology/shape of the traits vary enough between sexes to facilitate sex estimation that is accurate to sex assignment at birth. These secondary sexual characteristics (i.e., sex characteristics beyond the organs directly related to reproduction) are also influenced by other factors, including genetics, hormones, environmental impacts during growth and development, and even physical activity (depending on the sex trait). Keep in mind that when discussing the "sex" of an individual, a human made a designation about their appearance at birth. Whenever possible, we will denote this as sex assigned at birth (or assigned sex) to differentiate it from the biology of sex (with all variations contained therein). Regardless, the skeletal traits being interpreted by anthropologists when estimating sex are based on various biological processes and stem from the typical chromosomal sex differences and hormones. Thus, when a forensic or biological anthropologist is estimating the sex of an individual from their skeleton, they are estimating their assigned *biological sex*, and as we will see, this may or may not conform to their gender.

Variations in sexual development

The biology of sex, however, is not as binary as the male/female typology implies, as described above and in the dictionary definitions. The development of biological sex is complex and mutable. Many individuals are born with differences or combinations of elements that define stereotypical male or female biological sex—from chromosomal arrangements to genital morphology (Blackless et al., 2000). The terms "intersex" and "differences of sex differentiation" (DSD) may be used to describe a person (or community) with a condition that affects the process of sex differentiation (reproductive and genitourinary) during embryological development.

Published estimates of intersex/DSD birth frequencies differ widely depending on the source and are attributed to numerous factors, including a lack of agreement in what exactly constitutes an intersex/DSD condition (e.g., some conditions are apparent at birth, while others may be completely unrecognized throughout an individual's life) and simply poor sampling. However, a recent analysis suggests at least 1:1000 individuals have some form of chromosomal, hormonal, gonadal, or anatomical development that is atypical, which influence embryological sex development and the emergence of secondary sex characteristics (Ostrer, 2014). Intersex conditions have a range of impacts on structures associated with sex anatomy and morphology. To contextualize the continuum of biological sex, we can explore intersex conditions using examples touching on three primary concepts of biological sex—the chromosomal attributions of sex, sex assignment based on genital morphology and reproductive organs, and the hormonal basis of sex assignments. It should be noted that there is individual variation in the biological presentation of every condition, and biological presentation does not necessarily reflect identity, as an individual's identity develops socially (Fisher et al., 2016). These examples by no means encompass the full range of variations possible for biological sex (that's a book in itself), but are presented as an introduction to, and to emphasize, the complexity even within biological sex categorization.

Klinefelter syndrome (KS), associated with a surfeit of X chromosomes in an XY arrangement, impacts physical and sometimes cognitive development processes. It is most often diagnosed in individuals with a 47, XXY chromosomal arrangement. Estimates of frequency vary, but average around 1 in 1000 live births (Blackless et al., 2000). Individuals with KS produce low levels of testosterone (Simpson et al., 2003). This can impact aspects of sex development and maturation, leading to a variety of phenotypes, including incompletely descended testes, incomplete pubertal development, hypogonadism (small testes), fertility difficulties, unusually tall adult stature, gynecomastia (breast development), micropenis, and hypospadias (urethral orifice along the underside of the penis) (Simpson et al., 2003). Klinefelter individuals are typically raised, identify, and present as males although this is not always the case (Kreukels et al., 2018).

Variations in sex development can also occur with stereotypical chromosomal arrangement. A condition where cell receptors do not respond to androgens, which direct typical male sex development, is referred to as androgen insensitivity syndrome (AIS), and nonresponsiveness may be partial or complete. An individual born with complete androgen insensitivity (CAIS) is phenotypically female despite a 46, XY karyotype. The process of sex development produces a range of stereotypical female pelvic and genital structures, without the development of a uterus, cervix, or uterine (fallopian) tubes, and often with incompletely descended testes. This presentation is due to the role of androgens in the process of genitourinary and reproductive organ development. Individuals with CAIS are typically assigned female sex at birth, given the common presence of female external genitalia, and often present and identify as female (Fisher et al., 2016). Androgens, testosterone, and estrogen play an important role in bone turnover and skeletal sexual dimorphism. Because of the differential bone response to androgens and estrogen across anatomical locations (e.g., periosteal vs. endosteal deposition, growth plate, trabecular bone, etc.), androgen insensitivity has been linked with several bone presentations, including reduced vertebral bone mineral density (Danilovic et al., 2007; Marcus et al., 2000; Wiren, 2005) and reduced periosteal deposition during growth (Almeida et al., 2017). A case study of an adult woman with CAIS suggests that, prior to estrogen replacement therapy, long bone periosteal and cortical cross-sectional dimensions may measure intermediate between female and male ranges (Taes et al., 2009). Without estrogen replacement therapy during puberty, individuals with CAIS are generally tall relative to the US average height for females (Han, Goswami, Trikudanathan, Creighton, & Conway, 2008). Because estrogen stimulates epiphyseal fusion, tall statures are likely related to a prolonged skeletal growth period.

Endogenous and exogenous hormone environments play a key role in embryological development and throughout life. Sex development and maturation can be affected by hormone production as well as cell responsiveness (as in androgen insensitivity). With congenital adrenal hypoplasia (CAH), an enzymatic deficiency (most commonly 21-hydroxylase) limits the function of the adrenal gland, often leading to an insufficient production of the hormones cortisol and aldosterone. Instead, the adrenal gland produces high levels of androgens (Witchel & Azziz, 2011). There is a continuum of forms of CAH that vary in frequency, ranging from classic form called salt-wasting CAH, which causes a life-threatening imbalance in fluids and electrolytes when not managed medically, to nonclassic CAH, also known as late-onset CAH (Pang, 2003; Speiser et al., 2000). The frequency of CAH depends on form and varies by population with estimates ranging from 1:14,000 births in the United States and Canada to 1:282 births among the Yupik in Alaska (Blackless et al., 2000; Pang, 2003; Speiser et al., 1985). Throughout intrauterine development and postnatally, the adrenal glands of individuals with nonclassic CAH produce typical levels of cortisol and aldosterone, but also produce very high levels of androgens, associated with virilization, and is the most common cause of hyperandrogenism (Merke & Bornstein, 2005). While the condition of CAH is not necessarily tied to biological sex, it is often diagnosed in babies with XX chromosome arrangements because the effects are more apparent, given ambiguous genitalia often develops even when reproductive organs develop stereotypically. The reproductive organs and external genitalia of XY babies with CAH usually develop typically. All individuals with CAH will experience precocious (early onset) puberty without hormone management. Early growth spurts with a rapid linear growth establishes a child as tall for their age. However, the process of epiphyseal fusion begins early and limits further growth, leading to shorter-than-expected adult statures (Speiser et al., 2000; Witchel & Azziz, 2011).

These represent only a few examples of variations in biological sex. With such conditions, it is not always clear what biological sex category individuals should be assigned to if we operate with a binary sex paradigm. Is it based on chromosomes? Is it based on reproductive organs? What if there are a mixture of reproductive organs present? Is it based on external genitalia? External genitalia can be ambiguous. There is no standardized answer, and many times it comes down to the sex assigned to an individual at birth by medical practitioners—which are then documented on birth certificates. Until recently, birth certificates only had female or male options (discussed further below). Obviously, the definitions provided by *Merriam-Webster* and other dictionaries are overly simplified, creating a need for terms like "intersex" and "disorders/differences of sex differentiation." Definitions may depend upon who is making the distinction and for what goal, and regardless of the criteria used, assigned biological sex may not correlate with a person's gender or self-identity. See the following case study as an example.

Case study: Defining male and female in athletics

In humans, it is presumed that stereotypical sexual dimorphism confers males a physical advantage in size, strength, and speed. Historically, female participation in athletics was restricted over the misguided fear that it would damage delicate organ systems. As far-fetched as this idea seems now, these ideas shape how sports are seen and practiced today. Most sports are grouped by sex, meaning that athletic organizations must have some definition of sex by which to designate people into male and female athletics (incorrectly termed "gender tests"). In the past, the Olympics commission examined genitalia of female athletes to prohibit male athletes from disguising themselves as females to gain an advantage for the metaling podium (Pieper, 2016). While there are few examples of such unethical behavior, these "gender tests" or "sex verification" tests have mostly identified and outed athletes with various intersex conditions, spurring controversy around the place of intersex individuals in the Olympics and athletics in general.

Recently, this issue has gained attention throughout international sports and in the media. Since the 1990s, it has been acknowledged that genital examinations and chromosome testing are problematic biomarkers for binary sex distinctions (as described above) and that there are no scientifically accepted criteria to absolutely sort male from female (Karkazis & Jordan-Young, 2015). Because testosterone is typically associated with maleness and the physical advantages gained through steroid use, the International Association of Athletics Federation (IAAF) and the International Olympic Committee (IOC) began testing testosterone levels to use as the deciding factor for eligibility in women's international competitions when a female athlete's sex is questioned. IAAF had a limit of 10 nmol/L for eligibility as females in international competition due to hyperandrogenism (high natural production of testosterone in females). However, like other measures of dimorphism, although the mean values differ, the ranges for naturally produced testosterone in stereotypical male and female elite athletes overlap, and many differ from normal reference profiles (Healy, Gibney, Pentecost, Wheeler, & Sonksen, 2014). It has been suggested that many female athletes have testosterone levels that are naturally higher than the nonathletic female population (Cook, Crewther, & Smith, 2012). While testosterone is associated with increased capability of power, speed, and aggression, testosterone is not the sole factor that confers a competitive advantage or distinguishes male and female athletes. This arbitrary limit was used as though it is a hard biological distinction, indicating that those with levels above are not actually female and have an "unfair advantage," but there had been no scientific studies to support this. After a professional sprinter from India, Dutee Chad, successfully appealed to the Court of Arbitration for Sport (CAS) in 2015 to reverse a ruling of ineligibility due to hyperandrogenism, the IAAF commissioned a study to determine the effect of testosterone levels on performance in track-and-field events (Bermon et al., 2014). Multiple intersex (CAH and CAIS) athletes were competing in middle-distance races, and their data were eliminated from the study, establishing a limited range a priori. These researchers found that there was not an effect of testosterone levels on performance in female track-and-field athletes except in middle-distance events (400 m to 1 mile), pole vaulting, and hammer throw (Bermon & Garnier, 2017). However, no other body composition or cultural traits (e.g., limb length, lean body mass, VO2 max, socioeconomic background) were explored as factors conferring an advantage in competition. The IAAF drew a new 5 nmol/L limit for middle-distance races only ("Restricted Events" bemusingly leaving out pole vaulting and hammer throw) in international competitions based on this study (IAAF introduces new eligibility regulations for female classification, 2018). Female athletes with testosterone levels exceeding the new limit wishing to compete internationally in "Restricted Events" can either medically alter her natural body composition or change her specialization to short or long distances. She may also compete against male athletes, in intersex-only races (if they are offered), or limit her career to national competitions—assuming her country allows her to compete.

In sports, can a bright-line be drawn to define sex based upon an arbitrary concentration level of a single hormone? Should athletes who are born with these natural variants be punished for their genetics? How is this reconciled with other natural physical, genetic, social, and financial "advantages"? Someone born with the genetics for exceptionally tall stature obviously has an advantage when it comes to certain sports (e.g., basketball), yet in that case, we consider it just natural variation. What is the impact of economic access to intensive training and coaching on competitive performance? Why is it different when it comes to intersex individuals or people with naturally higher testosterone levels?

Defining gender

Unlike biological sex, gender is associated with social roles and constructs, particularly in anthropology. Gender is not biologically determined, and gender identity does not always match expectations associated with assigned biological sex. Definitions and terms are continuously developing, particularly among the LGBTQQIAP2 community (an initialization standing for Lesbian, Gay, Bi-sexual, Transgender, Queer, Questioning, Intersex, Asexual, Pansexual, and Two-Spirit; often abbreviated using a combination of initials).

Currently in the United States, popular culture and psychologists understand gender as a nonbinary construct (American Psychological Association, 2015). However, there is still a strong cultural and legal relationship to a urogenital or biological determination of gender (Westbrook & Schilt, 2014)—for example, in public accommodations, sports, and sex-segregated locations. With increasing awareness of nonbinary and gender diverse concepts, ideas of gender identity, expression, and sexual orientation are expanding and becoming more widely accepted. As this happens, people are responding by broadening the lexicon to define and redefine these ideas. In defining new terminology in this chapter, it is important to discuss them in the context of gender *identity* and gender *expression*. These concepts and definitions are complex, deeply personal, and probably more limited than the actual variation in individuals' experiences. Organizations like Lambda Legal and GLAAD (Gay and Lesbian Alliance Against Defamation) have resources available to stay up-to-date with new terms and conventions (GLAAD, 2016).

Gender identity is one's psychological sense of gender—how a person understands themselves. The spectrum of gender identity is culturally specific and reflects inherent cultural expectations and traits. In many cultures, binary gender identities of man and woman tend to conform, generally speaking, to binary traits and expectations associated with a binary biological sex. However, there is a broadening understanding that gender identity is much more complex than the binary norms. Terms reflecting the complexity of gender identity, which are extensive, continuously adapting, and nuanced, are becoming more common and accepted. In modern culture in the United States, a person whose sense of self that does not conform to binary expectations of gender roles may describe themselves through a mosaic of evolving identity terms, including gender nonconforming, queer, genderqueer, gender fluid, and nonbinary. A person whose gender identity does not conform to expectations associated with their sex assigned (or presumed) at birth could identify as transgender (see Chapter 20 of this volume). It is important to consider that concepts of gender identities are culturally specific. For example, Two-Spirit is a sacred and specific term used by many gender diverse Indigenous North American and First Nations individuals to reference gender diversity in identity, spiritual, and social roles (Driskill, Finley, Gilley, & Morgensen, 2011). The concept of Two-Spirit has a long and integrated history among many tribes in North America, and each tribal language may also have a traditional term for the identity.

The word "queer" can be an uncomfortable word for many people to say and hear given its history as a slur. However, it has been reclaimed that, in addition to representing a personal gender identity and operating as a sex- and gender-neutral term for sexual orientation (as opposed to gay or lesbian), it is often used as an in-group shorthand for the LGBTQ+ community (e.g., "the queer community"). While the word "queer" has been reclaimed, it is important to keep in mind that its derogatory history is intrinsic to the term and many in the community still feel that history, particularly when used by those who do not identify as LGBTQ+.

How one shows their gender identity is referred to as gender *expression* (or presentation). Expressions of gender are tied in with cultural expectations and performed through a variety of factors such as clothes/grooming, mannerisms, speech patterns, behavior, and pronouns (e.g., she/her, he/him, they/them). In western culture, expressions of feminine, masculine, and androgynous are commonly understood and the traits are specific and generally recognizable. Recently, gender expression has also been adapting and can include a mosaic of nonbinary expressions. The term "cisgender" is often used to describe someone whose gender identity and expression mirror their sex assigned at birth. For example, a cisgender male refers to a person who was assigned male sex at birth and they identify and express themselves in a way that conforms with cultural expectations associated with maleness. A person's expression is not necessarily static and may be performed differently depending upon a variety of factors, including environmental context, comfort, and safety concerns.

The term transgender is something of an umbrella term and identity that indicates a person has a gender identity that does not conform to the sex they were assigned at birth. This means something different for everyone. Although being a transgender (or trans for short) is often associated with having a gender identity that is "opposite" from the sex assigned at birth, it is not necessarily a binary experience. Being trans is a long process that may involve transitioning socially, medically, and legally. There are as many transition goals as there are trans people. The social transition process ranges from coming out to oneself and other people as trans to living and expressing your gender identity in all aspects of your life. Not everyone wants or is able to have all possible sex reassignment (also known as gender-affirming) surgeries. Similarly, hormone therapy is not a universal part of transitioning. Legal transitions are important milestones in a transition. Legal transitions involve changing assigned sex markers and legal names on identification documents (including driver's licenses, state-issued identification cards, birth certificates, social security cards, passports, etc.—discussed further below). For many people, parts of these processes are unattainable. There are numerous barriers to all these aspects of transitioning, which may include social or familial (e.g., hostile environment or culture, unsupportive family), financial (access to medical care, access to legal processes—discussed below), medical care (e.g., financial, unsupportive/inexperienced/hostile medical practitioners), and institutional (e.g., university databases) barriers. This means that while goals for transitioning may vary, so do the attainability and timing of transition processes. There is no checklist that can be applied to define a person's transition process.

The terms transgender and transsexual are commonly confused and conflated. This probably follows the confusion over sex and gender because transgender and transsexual are sometimes used synonymously. This is partially because the term "transsexual" is understood in different ways by different people. Transsexual is an older term, from the mid-1900s, originally used by medical and psychological professionals to describe a transgender person before concepts of gender were linked with self-identity in these fields. For some, transsexual holds the same meaning as transgender. For others, transsexual is used to specifically describe a (transgender) person who has changed or seeks to change their body through medical processes. The key difference in current usage is that transgender is an umbrella term, while many transgender people do not identify as transsexual whether they are transitioning medically or not.

Sexual orientation is not necessarily related to gender identity or expression. It is an indication of the people we are drawn to or attracted to. It harkens back to our and other's identity, expression, and sex, but is neither wholly dependent nor independent of those factors. Sexual orientation operates on a continuum. It is commonly thought to range from homosexual (same sex attracted: lesbian/gay) ↔ bisexual ↔ heterosexual (opposite sex attracted: straight). As with broadening understandings of gender identity and expression, these descriptions of attraction do not encompass all of the current understandings of sexual orientation. For example, pansexual describes a sexual orientation where there is attraction to a broad range of gender identities and expressions regardless of presumed sex. People also identify with various descriptions of their interest in sexual (sexual/asexual) and/or romantic relationships (romantic/aromantic). People may become confused using "identify" when describing sexual orientation and associate it with gender identity. The context of use will indicate which is being referred to. For example, a person may identify as a woman (gender identity), have an androgynous or nonbinary gender expression, and identify as straight (orientation).

Why so much confusion?

There are a couple of different factors contributing to the confusion between the terms gender and sex. The first, and the major takeaway from the sections above, is that both sex and gender are much more complex than they have been portrayed historically. Complexity and a lack of full comprehension can lead to misuse of the terms. The word "sex" is also still commonly portrayed as a "dirty" word given that it also refers to sexual intercourse. Avoidance of the term sex in combination with naivety of their appropriate definitions may cause further confusion. For example, how many people in our culture have a "sex-reveal" party when they first find out the biological sex assigned to their unborn child? Instead, the term "gender" is used, as it has a less provocative connotation, even though the presence of reproductive organs is being used to make this determination. At the point a fetus is developing, the debate around whether there is a genetic foundation to gender, or at what age gender can be ascribed, is immaterial since a fetus cannot communicate their identity. The term gender is commonly preferred in popular media and conversations, and even some medical and scientific literature still use gender when referring to biological processes (Krieger, 2003). Such an exchange of the terms sex and gender suggests that they are synonymous, which they are not, and adds to the confusion.

Government documents can also add to the confusion. Many forms require individuals to mark themselves as either male or female and may variably use the terms sex or gender. Driver's licenses typically list an individual's "sex" as either "M" for male or "F" for female, based on an assigned sex from individual's birth certificate. Transgender individuals can change the assigned sex on their driver's licenses, but the process varies by state. Some states require a medical proof of various surgeries, court order, or an amended birth certificate. Others require certification from any one of a broader array of licensed professionals indicating that they have had clinical treatment for gender transition. Since 2017, a few states have dropped any certification requirements (Oregon, Washington DC, Nevada, Maine, Massachusetts, California), and some have added a third gender-neutral option (Oregon, Washington DC, Maine, California) to better serve intersex, transgender, and nonbinary individuals (Grinberg, 2017; ID Documents Center, 2018; Nevada DMV makes it easier to change gender on licenses, IDs, 2018). While most may not think twice about their documented sex on their license, being able to change this documentation to match their identity provides validation to transgender individuals, reducing anxiety and possible societal conflicts that can make a person a target for harassment, discrimination, or violence (see Tobia, 2017 for more details). It does bring up some semantic issues, however, as driver's licenses still note an individual's "sex," while allowing individuals to change their category to better reflect their identity may, in many cases, be reflecting an individual's gender (which may not conform to their assigned biological sex). There are many reasons why an individual's gender identity is more appropriate on driver's licenses, but as we will discuss in the next section, there are some important considerations for forensic anthropological searches for unidentified individuals.

Finally, it is important to note that the use of sex and gender terms in social and political contexts can also drive confusion, sometimes purposely. Some terms, for example, sex, may hold different legal rights or recognition than others. Terms may be misused to drive political campaigns and to spur reactions from audiences. For example, in October 2018, just prior to a tense election in the United States, the Department of Health and Human Services suggested to change the federal definition of gender, in relation to Title IX and gender discrimination, to be determined by the "immutable" characteristics of male or female sex as assigned on an original birth certificate (Green, Benner, & Pear, 2018). We know from the discussion above that this distinction is fraught with a misunderstanding of the biological *mutability* of sex at its foundation, the relationship between sex and gender, and—given the timing and wording—it had a lot more than a whiff of political motivation.

A quick look at *Meriam-Webster*'s online definitions of "gender," and the subsequent debate held in the comments section following the definitions, highlights some of the areas of confusion and discrimination around sex and gender (Gender, n.d.). The first definition of "gender" deals with grammar. The second definition defines gender as "(a) sex, the feminine *gender* and (b) the behavioral, cultural, or psychological traits typically associated with one sex." While the second part of the second definition begins to approach what we have defined here as gender (i.e., social traits), it still associates gender with sex and assumes that they align with one another. One user comment states "gender and sex are synonyms and always have been." Another user comments "trying to understand what gender is. One definition said, sex = male and female, gender = feminine and masculine, but that clearly can't be right. I've know[n] men who were very feminine who were still men and visa [sic] versa, so when someone defines their gender as female, for example, it means more tha[n] simply feminine." The comments regarding sex and gender—even on this online dictionary webpage, a place where people refer to for clarification on definitions—become heated debates at times (including name-calling). This exemplifies the confusion around the terms and how personal people take the use and misuse of the terms.

Implications in forensic and biological anthropology

Forensic anthropologists are often asked to estimate an individual's biological profile from a set of skeletal remains—the information from which investigators and law enforcement use to search their "missing persons" files in an attempt to identify the unknown decedent. Given that the forensic anthropologist is working only with the skeletal remains (which form from biological processes), the anthropologist is ultimately estimating the biological sex of the individual. Based on the individual's chromosomes and associated hormones, certain skeletal regions take on a more (biological) male or female form. For example, the shape of the pelvic inlet attributed to typical biological females is relatively wider to provide the potential for childbirth, which is associated with other morphological changes and traits in the pelvis. While there are gradients in skeletal traits with a degree of overlap between stereotypical assigned male and female forms, studies report agreement between sex estimates and sex (assigned at birth) in the high 90 percentiles when utilizing the pelvis. These skeletal traits, however, tell the forensic anthropologist nothing about an individual's gender identity, gender expression, or sexual orientation.

Because there is a continuum of trait expressions, several sex estimation methods involve scoring certain skeletal features on an ordinal scale, and then those numbers are subsequently applied to multivariate equations for sex estimation. For example, the Walker (2008) method entails scoring the expression of glabella/supraorbital ridge, supraorbital margin, mastoid process, nuchal crest, and the mental eminence on a scale of 1–5. For this method, a score of 1 represents a hyper-gracile morphology, and 5 a hyper-robust morphology. Sometimes the terms hyper-feminine or hyper-masculine are used by practitioners (following Acsádi & Nemeskéri, 1970) (e.g., Gómez-Valdés et al., 2012; Ramsthaler, Kreutz, & Verhoff, 2007). In actuality, these numbers represent where a single trait expression falls along the broad spectrum of trait expressions, in which male morphologies tend to group toward the higher end of the ordinal scale, and female morphologies toward the lower end. Frequently it is a measure of degree of robusticity/gracility. Such methods do not mean that a male individual cannot have a female trait expression (e.g., a score of 1 or 2). As noted previously, there are overlaps between male and female morphologies. Furthermore, anthropologists will look at a whole suite of traits available prior to making an estimate, knowing that it is common for individuals to show variation in expressions among the traits. Note also that an individual who displays a mixture of trait expressions (e.g., a 4 for mastoid process, but a 1 for mental eminence) is not indicative that the individual is transgender, intersex, or more or less female/feminine or male/masculine in terms of sex or gender. These skeletal trait expressions are not connected to gender expression and tells the practitioner nothing about how the individual self-identifies, expresses, or behaves.

The use of gender vs. sex can be an issue in forensic cases involving transgender individuals—for example, if agencies are searching for a missing individual assigned male at birth, when the individual is a transgender woman whom everyone in their community knows as a female. This situation has received some media attention lately from a ProPublica article (Waldron & Schwencke, 2018). The act of calling a transgender individual by their previous name and sex is referred to as dead-naming. The article explains that when police use the previous (dead) assigned sex, name, and pronouns of a transgender individual, it is not only disrespectful but also solidifies a distrust of law enforcement within a community where the trust and relationship with law enforcement is already strained. This can impede their identification and slow down investigations into their deaths. The law enforcement agency approached in this article indicated that it is their policy to use the names and sex listed on the victim's state-issued identification. However, some law enforcement agencies do have policies to refer to victims as they identified. While relaxing the requirements to change documented sex on one's license may help resolve part of this conflict, it is crucial for law enforcement and investigative agencies to institute policies that respect a person's lived identity at the time of death as there are many barriers to transitioning legally.

What if law enforcement and coroners/medical examiners have an unidentified body found without any clothing or associated material artifacts? They may have no indication of a person's gender. They can only assign sex based on genitalia and reproductive organs. What about a scenario where the remains were completely skeletonized and all they had was the estimated sex from the forensic anthropologist. If the decedent is part of the trans community and people are aware of the decedent's transgender status and reported it to the investigators, an identification may still be made (the skeletal sex markers may not match the individual's lived gender expression). However, what if the decedent is not from that geographic area/community. Perhaps someone killed the decedent and then transported and disposed off the body in a different state. Would authorities searching "missing persons" databases for an individual with an assigned male sex at birth be able to connect the dots if the missing individual is a trans woman and listed in databases as a female? One idea is to have such databases include both sex and gender information, so that those searching to identify a decedent can use whatever information is available and increase the chances of identification. From the outside perspective of someone merely wishing to help identify a decedent, bring closure to their family and community, and help bring justice for their death, this may seem like a good idea; however, some unintended consequences must be considered. Again, individuals may see any documentation of their dead sex (or their friend's dead assigned sex) as disrespectful, given the long and torturous journey to realizing changes to documentation. Any indication of gender diversity in a database could also target individuals or their peers for harassment, violence, and discrimination. Yet, without such information, a decedent may remain unidentified and any persons responsible for their deaths free at large. It is a challenging dilemma. Public accessibility of such databases needs to be considered. Certain databases, such as NAMUS (https://www.namus.gov/), allow for case managers to decide which, if any, of the information should be available to the public, while the remaining information can only be accessed by validated authorities pertinent to such investigations. At present, NAMUS only has a single "sex" category and permits only "male," "female," or "unsure" responses, but additional searchable information can be provided in "circumstances" text boxes (which again may be kept private and out of public view). Perhaps nonbinary responses for such categories could be included, similar to the gender-neutral option available on some state licenses, but again could face similar issues by outing an individual. The concepts of sex and gender are not uncomplicated, nor are considerations associated with them. These concepts and their official applications have real-world effects and are becoming increasingly important to address in the medicolegal field. Ultimately, when it comes to such decisions regarding the advancement of identification processes, it is vital that discussions are held with individuals within the affected communities so that all perspectives are included in the decision-making processes and everyone is aware of all possible implications. Communication and collaboration are vital for success, especially when both parties have a common goal—to help the decedent.

We should also note at this point that just because the biological traits of a skeleton may be pointing toward a particular sex estimate (say female) and material evidence at the scene suggests the opposite sex (e.g., men's clothing), it should not be assumed that the individual is a trans man or any other diverse identity. First, as mentioned above, sexually dimorphic traits in the skeleton form a continuum, and there is overlap in traits associated with male and female skeletal features. It is possible that the skeletal estimate may be incorrect. Even if the individual was assigned female sex at birth, the presence of male clothing does not necessarily mean the individual identifies as transgender (or any other diverse identity). Remember gender identity and gender expression are different. The individual very well may be a cisgender female who happens to prefer or only has access to men's clothing. Even if an individual was cross-dressing at the time of their death, it should be considered that this may have been done in private, and investigators should act considerately with sensitive information when approaching friends and family of the decedent. The key is not to jump to conclusions, keep an open mind, and be sensitive in any further inquiries.

The impact of hormone therapy on sexually dimorphic skeletal traits has not been well documented. If a transgender individual undergoes hormone treatment postpuberty, once their skeleton has already fully developed, any skeletal changes may be minimal and may not have a major impact on skeletal sex estimation. With advances in medicine and cultural acceptance in recent years, some people are able to start transitioning medically and socially prior to puberty. Individuals may also choose to medically delay puberty to allow additional time to make decisions. As this population ages, we may need to address potential impacts on skeletal features used in sex estimation. Until then, there are some processes of medical transitions associated with hormone replacement therapy (HRT) and surgeries that may be apparent in skeletal material. Transgender men who have started hormone therapy and have increased skeletal muscle strength may have robust bones and muscle attachments associated with assigned male skeletal features. Transgender individuals may also choose to undergo bone-modifying surgeries (particularly in male-to-female transitions). Postcranially, these may include rib removal to narrow the waist or amputation of the fifth pedal digit to narrow the foot ("stiletto surgery"). Surgeries are also available to "feminize" facial features. Forehead reduction, rhinoplasty, genioplasty, and shaving of the mandibular angle are possible facial procedures in male-to-female surgeries (Altman, 2012; Buchanan, 2014). These procedures, unlike many other aspects of gender reassignment or affirming surgery, directly impact the skeletal features and may affect skeletal sex classifications. On the other hand, if evidence of these procedures or hormone treatment are observed, it could provide additional clues in the search for the identity of the individual.

The skeletal morphology of intersex/DSD individuals depends greatly on the condition. Although there are suggestions of clinically relevant skeletal impacts related to hormone replacement (e.g., in androgen insensitivity discussed earlier), specific pelvic and skeletal morphologies associated with skeletal sex estimation do not appear to have been addressed in the anthropological literature. Because even nonintersex individuals (stereotypical males and females) can display ambiguous traits, the presence of such traits is not indicative of an intersex condition. A forensic anthropologist will not be able to determine whether an individual was born with a DSD condition based on skeletal analyses. DNA analyses, however, can reveal chromosomal conditions.

Conclusion

Anthropological texts generally state that sex is biological and gender is cultural, and while this is true, classifications within each of these terms are complex, and a lack of understanding or misuse of sex and gender terms can have major implications on forensic anthropological analyses. Biological sex is commonly portrayed as a dichotomy, despite the fact that a number of chromosomal, genetic, and hormonal variants exist (e.g., intersex individuals) that do not neatly prescribe biological male or biological female category. This does not mean that forensic anthropologists cannot or should not estimate sex (as accuracy rates are in the high 90 percentiles and contribute significantly to identifications), but practitioners should be aware of these potential variations, as they may be identified through DNA analyses or investigative efforts and may result in discordance between the sex estimated from skeletal features and an individual's self-identity. Although the forensic anthropologist is estimating biological sex from the skeleton, it is still important to keep up-to-date on gender terms as their definitions continue to evolve and new terms arise. The forensic anthropologist is in a unique position where they are working to connect biological information (from the skeleton) to sociocultural information (how a person self-identifies or is identified by others). To make those connections, it is integral to be familiar with all relevant terms, both to increase identification successes and to be respectful to the deceased and their community.

Finally, please note that the material presented in this chapter primarily refers to cultural terms as used in the United States. Different geographic regions, cultures, or societies may have their own version of gender terms and definitions. It is important to be familiar with the culture in which you are working. It is also likely that, even as this book comes to print, some of the terms or definitions may change as we continue to become more open as a society and learn more about ourselves.

References

Acsádi, G., & Nemeskéri, J. (1970). *History of human life span and mortality*. Akadémiai Kiadó.

Almeida, M., Laurent, M. R., Dubois, V., Claessens, F., O'Brien, C. A., Bouillon, R., et al. (2017). Estrogens and androgens in skeletal physiology and pathophysiology. *Physiological Reviews*, *97*(1), 135–187. https://doi.org/10.1152/physrev.00033.2015.

Altman, K. (2012). Facial feminization surgery: Current state of the art. *International Journal of Oral and Maxillofacial Surgery*, *41*(8), 885–894. https://doi.org/10.1016/j.ijom.2012.04.024.

American Psychological Association. (2015). Guidelines for psychological practice with transgender and gender nonconforming people. *American Psychologist*, *70*(9), 832–864. https://doi.org/10.1037/a0039906.

Bermon, S., & Garnier, P. Y. (2017). Serum androgen levels and their relation to performance in track and field: Mass spectrometry results from 2127 observations in male and female elite athletes. *British Journal of Sports Medicine*, *51*(17), 1309–1314. https://doi.org/10.1136/bjsports-2017-097792.

Bermon, S., Garnier, P. Y., Hirschberg, A. L., Robinson, N., Giraud, S., Nicoli, R., et al. (2014). Serum androgen levels in elite female athletes. *The Journal of Clinical Endocrinology & Metabolism*, *99*(11), 4328–4335. https://doi.org/10.1210/jc.2014-1391.

Blackless, M., Charuvastra, A., Derryck, A., Fausto-Sterling, A., Lauzanne, K., & Lee, E. (2000). How sexually dimorphic are we? Review and synthesis. *American Journal of Human Biology*, *12*, 151–166. https://doi.org/10.1002/(SICI)1520-6300(200003/04)12.

Buchanan, S. (2014). *Bone modification in male to female transgender surgeries: Considerations for the forensic anthropologist*. LSU Master's Thesis. 1290. Retrieved from: https://digitalcommons.lsu.edu/gradschool_theses/1290.

Cook, C. J., Crewther, B. T., & Smith, A. A. (2012). Comparison of baseline free testosterone and cortisol concentrations between elite and non-elite female athletes. *American Journal of Human Biology*, *24*(6), 856–858. https://doi.org/10.1002/ajhb.22302.

Danilovic, D. L. S., Correa, P. H. S., Costa, E. M. F., Melo, K. F. S., Mendonca, B. B., & Arnhold, I. J. P. (2007). Height and bone mineral density in androgen insensitivity syndrome with mutations in the androgen receptor gene. *Osteoporosis International*, *18*(3), 369–374. https://doi.org/10.1007/s00198-006-0243-6.

Driskill, Q. -L., Finley, C., Gilley, B. J., & Morgensen, S. L. (2011). Introduction. In Q. -L. Driskill, C. Finley, B. J. Gilley, & S. L. Morgensen (Eds.), *Queer indigenous studies: Critical interventions in theory, politics, and literature* (pp. 1–28). Tucson: University of Arizona Press.

Fisher, A. D., Ristori, J., Fanni, E., Castellini, G., Forti, G., & Maggi, M. (2016). Gender identity, gender assignment and reassignment in individuals with disorders of sex development: a major of dilemma. *Journal of Endocrinological Investigation*, *39*(11), 1207–1224. https://doi.org/10.1007/s40618-016-0482-0.

Gender (n.d.). Retrieved February 15, 2018 from: https://www.merriam-webster.com/dictionary/gender.

GLAAD. (2016). *Media reference guide—10th edition*. Retrieved from: http://www.glaad.org/sites/default/files/GLAAD-Media-Reference-Guide-Tenth-Edition.pdf.

Gómez-Valdés, J. A., Quinto-Sánchez, M., Garmendia, A. M., Veleminska, J., Sánchez-Mejorada, G., & Bruzek, J. (2012). Comparison of methods to determine sex by evaluating the greater sciatic notch: Visual, angular and geometric morphometrics. *Forensic Science International*, *221*(1–3), 156-e1.

Green, E. L., Benner, K., & Pear, R. (2018, October). *Transgender. In Could be defined out of existence under trump administration*: New York Times. Retrieved from: https://www.nytimes.com/2018/10/21/us/politics/transgender-trump-administration-sex-definition.html?module=inline.

Grinberg, E. (2017, June). *You can now get a gender neutral driver's license in DC*. CNN. Retrieved from: https://www.cnn.com/2017/06/27/health/washington-gender-neutral-drivers-license/index.html.

Han, T. S., Goswami, D., Trikudanathan, S., Creighton, S. M., & Conway, G. S. (2008). Comparison of bone mineral density and body proportions between women with complete androgen insensitivity syndrome and women with gonadal dysgenesis. *European Journal of Endocrinology*, *159*(2), 179–185. https://doi.org/10.1530/EJE-08-0166.

Healy, M., Gibney, C., Pentecost, M., Wheeler, M., & Sonksen, P. (2014). Endocrine profiles in 693 elite athletes in the postcompetition setting. *Clinical Endocrinology*, *81*(2), 294–305. https://doi.org/10.1111/cen.12445.

IAAF introduces new eligibility regulations for female classification. (2018). Retrieved from: https://www.iaaf.org/news/press-release/eligibility-regulations-for-female-classifica.

ID Documents Center. (2018). Retrieved from: https://transequality.org/documents.

Karkazis, K., & Jordan-Young, R. (2015). Science and society. Debating a testosterone "sex gap". *Science*, *348*(6237), 858–860. https://doi.org/10.1126/science.aab1057.

Kreukels, B. P., Köhler, B., Nordenström, A., Roehle, R., Thyen, U., Bouvattier, C., et al. (2018). Gender dysphoria and gender change in disorders of sex development/intersex conditions: Results from the dsd-LIFE Study. *The Journal of Sexual Medicine*, *15*, 777–785.

Krieger, N. (2003). Genders, sexes, and health: What are the connections—And why does it matter? *International Journal of Epidemiology*, *32*(4), 652–657. https://doi.org/10.1093/ije/dyg156.

Marcus, R., Leary, D., Schneider, D. L., Shane, E., Favus, M., & Quigley, C. A. (2000). The contribution of testosterone to skeletal development and maintenance: Lessons from the androgen insensitivity syndrome. *Journal of Clinical Endocrinology and Metabolism*, *85*(3), 1032–1037. https://doi.org/10.1210/jcem.85.3.6428.

Merke, D. P., & Bornstein, S. R. (2005). Congenital adrenal hyperplasia. *Lancet*, *365*, 2125–2136. https://doi.org/10.1056/NEJMra021561.

Nevada DMV makes it easier to change gender on licenses, IDs. (2018, June). *Reno Gazette journal*. Retrieved from: https://www.rgj.com/story/news/2018/06/18/dmv-makes-easier-change-gender-licenses-ids/710081002/.

Ostrer, H. (2014). Disorders of sex development (DSDs): An update. *Journal of Clinical Endocrinology and Metabolism*, *99*(5), 1503–1509. https://doi.org/10.1210/jc.2013-3690.

Pang, S. (2003). Newborn screening for congenital adrenal hyperplasia. *Pediatric Annals*, *32*, 516–523. https://doi.org/10.3928/0090-4481-20030801-09.

Pieper, L. P. (2016). *Sex testing: Gender policing in women's sports*. Champaign: University of Illinois Press.

Ramsthaler, F., Kreutz, K., & Verhoff, M. A. (2007). Accuracy of metric sex analysis of skeletal remains using Fordisc® based on a recent skull collection. *International Journal of Legal Medicine*, *121*(6), 477–482.

Sex (n.d.). Retrieved June 29, 2018 from: https://www.merriam-webster.com/dictionary/sex.

Simpson, J. L., de la Cruz, F., Swerdloff, R. S., Samango-Sprouse, C., Skakkebaek, N. E., Graham, J. M., et al. (2003). Klinefelter syndrome: Expanding the phenotype and identifying new research directions. *Genetics in Medicine*, *5*(6), 460–468. https://doi.org/10.1097/01.GIM.0000095626.54201.D0.

Speiser, P. W., Dupont, B. O., Rubinstein, P., Piazza, A., Kastelan, A., & New, M. I. (1985). High frequency of nonclassical steroid 21-hydroxylase deficiency. *Obstetrical and Gynecological Survey*, *41*(4), 244–245. https://doi.org/10.1097/00006254-198604000-00017.

Speiser, P. W., Knochenhauer, E. S., Dewailly, D., Fruzzetti, F., Marcondes, J. A. M., & Azziz, R. (2000). A multicenter study of women with nonclassical congenital adrenal hyperplasia: Relationship between genotype and phenotype. *Molecular Genetics and Metabolism*, *71*(3), 527–534. https://doi.org/10.1006/mgme.2000.3036.

Taes, Y., Lapauw, B., Vandewalle, S., Zmierczak, H., Goemaere, S., Vanderschueren, D., et al. (2009). Estrogen-specific action on bone geometry and volumetric bone density: Longitudinal observations in an adult with complete androgen insensitivity. *Bone*, *45*(2), 392–397. https://doi.org/10.1016/j.bone.2009.04.198.

Tobia, J. (2017, June). *Why transgender people suffer from 'ID Anxiety'*. TIME. Retrieved from: http://time.com/4817285/transgender-id-anxiety/.

Waldron, L., & Schwencke, K. (2018, August). *Deadnamed*. ProPublica. Retrieved from: https://www.propublica.org/article/deadnamed-transgender-black-women-murders-jacksonville-police-investigation/amp?__twitter_impression=true.

Walker, P. L. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, *136*(1), 39–50. https://doi.org/10.2217/bmm.13.83.

Westbrook, L., & Schilt, K. (2014). Doing gender, determining gender: transgender people, gender panics, and the maintenance of the sex/gender/sexuality system. *Gender and Society*, *28*(1), 32–57. https://doi.org/10.1177/0891243213503203.

Wiren, K. M. (2005). Androgens and bone growth: It's location, location, location. *Current Opinion in Pharmacology*, *5*(6 SPEC. ISSUE), 626–632. https://doi.org/10.1016/j.coph.2005.06.003.

Witchel, S. F., & Azziz, R. (2011). Congenital adrenal hyperplasia. *Journal of Pediatric and Adolescent Gynecology*, *24*(3), 116–126. https://doi.org/10.1016/j.jpag.2010.10.001.

CHAPTER 5

Effect of sex misclassification on the skeletal biological profile

Diana L. Messer[a], **Sara M. Getz**[b]
[a]Department of Applied Forensic Sciences, Mercyhurst University, Erie, PA, United States
[b]Department of Anthropology, Idaho State University, Pocatello, ID, United States

Introduction

The estimation of sex is considered one of the most critical components in the construction of a skeletal biological profile. It is often addressed first in any skeletal analysis as many methods for other parameters of the biological profile are either sex-specific or may be interpreted differently based on the sex estimate (Christensen, Passalacqua, & Bartelink, 2014). Sex estimates are also critical from a medicolegal standpoint because this information effectively eliminates a large fraction of potential matches from consideration.

In cases when the majority of a skeleton is present, metric and morphological sex estimation methods based on the skull and pelvis are the most common; however, over the past century, methods for sex estimation have been developed for use with many bones of the adult skeleton, and metric sex estimation from the long bones is gaining popularity (Spradley & Jantz, 2011). The various techniques for sex estimation are discussed throughout this book.

Even for methods with high accuracy, the overlapping nature of human sexual dimorphism means that the possibility of misclassification is always present. This is particularly true if portions of the skeleton are damaged, fragmented, or not available for analysis. This chapter briefly discusses the theoretical potential for error in the biological profile resulting from misclassification of sex and presents supporting data from a small sample of positively identified forensic cases from the Department of Applied Forensic Sciences at Mercyhurst University. The practical application of these discussions is explored in a case study.

Sex and age estimation

According to a survey of 145 forensic anthropology practitioners (Garvin & Passalacqua, 2012), adult age-at-death is most frequently estimated using features of the pubic symphysis (Brooks & Suchey, 1990; Gilbert & McKern, 1973; Katz & Suchey, 1986, 1989; McKern & Stewart, 1957; Suchey, Brooks, & Katz, 1988; Todd, 1920, 1921), sternal end

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00005-5

of the fourth rib (İşcan & Loth, 1986a, 1986b; İşcan, Loth, & Wright, 1984a, 1984b, 1985, 1987), iliac auricular surfaces (Buckberry & Chamberlain, 2002; Lovejoy, Meindl, Pryzbeck, & Mensforth, 1985; Osborne, Simmons, & Nawrocki, 2004), and cranial sutures (Meindl & Lovejoy, 1985; Nawrocki, 1998). These commonly used approaches provide point estimates of age and associated confidence intervals based on the distribution of traits in the reference sample(s) from which each method was developed. With the exception of methods for the auricular surface and some cranial suture techniques, these methods generate sex-specific age estimates. Although age-specific differences in the progression of traits in each anatomical region may exist, the way in which age ranges are calculated with traits or phases within most methods means that the difference between males and females is directly, and perhaps primarily, related to differences in male and female age distributions in the reference samples. This means that the magnitude of error associated with an incorrect sex estimate is primarily determined by the methods chosen and how they are combined into a final age estimation.

The six-phase Suchey-Brooks method (Brooks & Suchey, 1990; Katz & Suchey, 1986, 1989; Suchey et al., 1988) is one of the—if not the most—commonly used methods for adult age estimation in both forensic and archaeological settings (Garvin & Passalacqua, 2012; Falys & Lewis, 2011). Each phase is described with a combination of features associated with a mean and age interval calculated from a sample of 1012 pubic symphyses (739 males, 273 females) collected during autopsies in Los Angeles, CA (Brooks & Suchey, 1990). As shown in Table 1, the mean values for each of the six stages are similar but not identical for males and females. Depending on the phase assigned to a particular individual, the difference between male and female means (point estimates) ranges from between 0.9 and 3.0 years, which, in practical terms, would have a negligible effect on the estimate produced. Of slightly greater importance are differences in the width of approximate 95% confidence intervals (calculated from the mean and standard deviation provided in Brooks & Suchey, 1990; Table 1) associated with each

Table 1 Comparison of age estimates (in years) produced by the Suchey-Brooks pubic symphysis method.[a]

	Male		Female		Difference (F–M)	
Phase	**Mean**	**Interval (±2 SD)**	**Mean**	**Interval (±2 SD)**	**Mean**	**Interval length**
I	18.5	14.3–22.7	19.4	14.2–24.6	0.9	2.0
II	23.4	16.2–30.6	25.0	15.2–34.8	1.6	5.2
III	28.7	15.7–41.7	30.7	14.5–46.9	2.0	6.4
IV	35.2	16.4–54.0	38.2	16.4–60.0	3.0	6.0
V	45.6	24.8–66.4	48.1	18.9–77.3	2.5	16.8
VI	61.2	36.8–85.6	60.0	35.2–84.8	−1.2	0.8

[a] Phases and means from Brooks and Suchey (1990), Table 1.

phase. The age intervals are between 0.8 and 16.8 years wider in females, but are only greater than 6.5 years in phase 5. In any case, the substantial difference in that phase is essentially negated by the recommendation in the original publication that no upper bound be used for phases 5 and 6.

After the pubic symphysis, the auricular surface and sternal rib ends are reportedly used by an equal percentage of practitioners for age estimation in forensic settings (Garvin & Passalacqua, 2012). Unlike the pubic symphysis and sternal rib ends, the most commonly used methods for estimating age from the iliac auricular surfaces (Buckberry & Chamberlain, 2002; Lovejoy et al., 1985; Osborne et al., 2004) do not have sex-specific descriptions or age estimates. With the exception of apical changes that may be influenced by more significant preauricular sulcus formation in females (Lovejoy et al., 1985) and some more recent evidence indicating sex-specific differences in the appearance and progression of several textural traits (Igarashi, Uesu, Wakebe, & Kanazawa, 2005), the features and progression of morphological changes in this joint are assumed to generally be the same for both sexes throughout adulthood.

Since their development in the early 1980s, the methods produced by İşcan and colleagues (İşcan et al., 1984a, 1984b, 1985, 1987; İşcan & Loth, 1986a, 1986b) are most likely to be used to evaluate sternal rib ends. Comparing the suggested point estimates and intervals for males and females from İşcan et al. (1984a, 1985) reveals that point estimates for phases 1–8 differ by between 0.5 and 6.0 years. As with the pubic symphysis, of greater interest is the difference in the length and positioning of associated intervals. The lower and upper bounds of the intervals for males and females differ only between 0.7 and 5.4 years and 0.4 and 7.1 years, respectively; however, the intervals for most phases are nonoverlapping and systematically too narrow to represent biological reality. The ways in which practitioners deal with this issue and report estimates from these methods are not standardized; thus, the potential effect of misclassification of sex is unknown, but likely minimal given the high degree of concordance between male and female point estimates and intervals in most phases.

In recent years, a revision of İşcan and colleagues' methods—Hartnett (2010)—has gained popularity. Hartnett evaluated the original İşcan stages (İşcan et al., 1984a, 1984b, 1985) in a large, modern autopsy sample containing bilateral sternal rib ends from 419 males and 211 females, ranging in age from 18 to 99 years (Hartnett, 2010). After a careful evaluation and assessment, the ribs were sorted into seven categories, and the İşcan descriptions were modified for clarity with a new emphasis on bone quality and density. Although the new mean ages for many of the phases are significantly different than the corresponding İşcan stages for both males and females, the mean ages for phases 1–7 show only minor differences between the sexes (Fig. 1) (Hartnett, 2010). Six of the seven means differ by less than or equal to 1 year, with phase 6 exhibiting a difference of just over 4 years, potentially as a result of a nontrivial difference in sample size in this age group (M: $n=61$; F: $n=18$). Comparing the ranges for each phase, shown in Fig. 1,

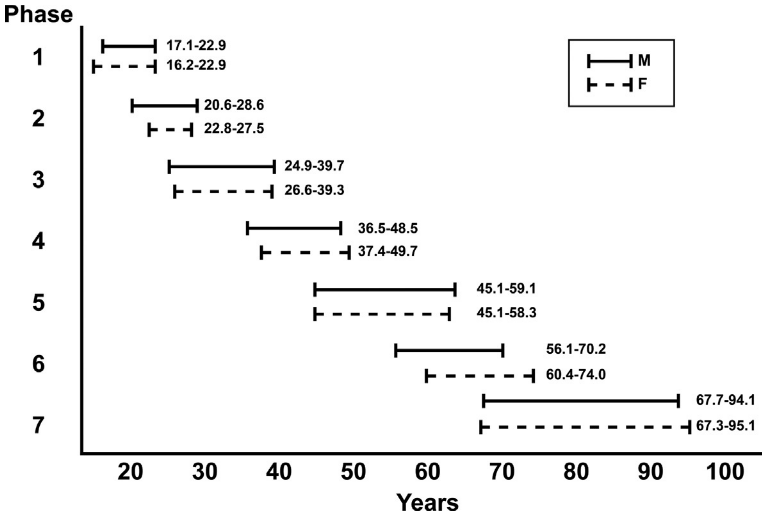


Fig. 1 Comparison of 95% confidence intervals from sternal rib end in phases 1–7 for males and females. *(Adapted from Hartnett, K. M. (2010). Analysis of age-at-death estimation using data from a new, modern autopsy sample. Part II. Sternal end of the fourth rib. Journal of Forensic Sciences, 55(5), 1152–1156, Tables 5 and 6).*

reveals that while differences between males and females are present and may potentially be biologically significant, they are not large enough to cause practically significant errors in age estimates in most cases.

Studies investigating cranial sutures do not consistently agree whether sex-specific equations are necessary. For example, Nawrocki (1998) provides sex-specific estimate equations, while Meindl and Lovejoy (1985) present age ranges that are independent of sex. In either case, over a century of research has failed to greatly alter the proposition about cranial sutures, put forth by Thomas Dwight (1890), that the progression of cranial sutures is too variable to be of significant aid to estimating age. Although they experience unidirectional change and can be scored with relatively low error, their progression is poorly related to age regardless of what constellation of features are scored or how they are analyzed (Cox, 2000; Hershkovitz et al., 1997; Milner & Boldsen, 2012). Thus, regardless of whether combined or sex-specific methods are used, age estimates are likely to be extremely wide, provide essentially the same information, and have little impact on the final estimate produced.

Although the most common features used in adult age estimation have been discussed here individually, rarely do practitioners use a single feature in isolation, and the ways in which they are combined to produce a final estimate are not standardized (Garvin & Passalacqua, 2012). Transition analysis (TA; Boldsen, Milner, Konigsberg, & Wood, 2002) is the only method currently available that statistically combines aspects of cranial sutures, pubic symphyses, and iliac auricular surfaces. In this method, probabilistic age information from independently scored components of each joint is combined to produce a maximum likelihood estimate of age with associated confidence intervals. The result is a probabilistically tailored maximum likelihood point estimate of age with a confidence interval for each individual based on the skeletal features present and consistency of age information provided by each trait. In contrast to the other techniques discussed, age estimates produced by the technique are not direct reflections of the age distribution of traits in the reference sample. As a result of the way in which age is calculated, the error introduced by an incorrect sex assessment will vary based on the age of the individual, the features present for analysis, and the collective suite of traits in all areas. Thus, the error introduced by an incorrect sex classification must be assessed on a case-by-case basis. Some examples from positively identified cases are presented later in this chapter.

As discussed, the magnitude of discrepancy between age estimates produced for the same individual, but assuming incorrect sex classifications, will ultimately depend on the specific methods chosen and the way in which they are combined. The average difference in point estimates and intervals for commonly used methods are relatively small, and multiple features are typically combined to generate a final estimate. Therefore, in most cases, misclassification of sex would likely result in a change in age estimate in the order of several years, as opposed to a decade or more. While this discrepancy may be particularly important in the youngest age categories or in research studies evaluating the aging process, it is unlikely to have significant practical effects for most adults in forensic contexts.

Sex and ancestry estimation

The estimation of ancestry is one aspect of the biological profile that is inextricably linked to sex. Sexual dimorphism varies between populations (see Chapter 17 of this volume) and has known, but not necessarily quantified, effects on the accuracy of skeletal sex estimation methods, particularly when metric analyses are used. In humans, understanding the relationship between sexual dimorphism and overall size requires the consideration of differing degrees and patterns of dimorphism both within and between populations.

Sex estimation using cranial and postcranial measurements is frequently analyzed using linear discriminant function analysis (LDFA), a statistical method used to determine which variables best discriminate between two or more groups (Jantz & Ousley, 2005, 2010). FORDISC, a statistical package and database that utilizes LDFA, has become standard in forensic contexts for estimating sex, ancestry, and stature (Cabo, Brewster, & Azpiazu, 2012; Dirkmaat, Cabo, Ousley, & Symes, 2008) (see Chapter 12 of this volume).

Variation in the degree of sexual dimorphism among populations introduces a potential source for error when estimating sex in both metric and morphological methods. In particular, the potential for error is emphasized when using LDFA, which magnifies the variation. In fact, the reliability of analyzing metric data with LDFA in FORDISC has been called into question for the evaluation of populations with differing levels of sexual dimorphism than what is typically found in American whites and blacks (Guyomarc'h & Bruzek, 2011; Spradley, Jantz, Robinson, & Peccerelli, 2008). Spradley et al. (2008) found that, using data from the Forensic Anthropology Data Bank, Hispanics were smaller postcranially than American whites, which leads to poor sex estimation, with too many males being classified as female. Similarly, when Guyomarc'h and Bruzek (2011) applied LDFA to the craniometric data from a modern Thai sample using FORDISC 3.1, they also reported poor sex classification accuracies.

Rather than a fundamental flaw of the statistical technique, these results are likely explained by the variation in sexual dimorphism between populations and, of course, by the fact that appropriate representative populations are not present in FORDISC 3.1. The degree of sexual dimorphism within the Hispanic and Thai populations tends to be significantly lower than the American white and black populations. Discriminant functions derived from white populations, which have a high level of sexual dimorphism, are likely to perform poorly for populations with a low level of dimorphism. Ultimately, it is likely that skeletal variation between ancestral groups should be taken into account when estimating sex by means of LDFA, particularly with the craniometric data.

While evidence clearly suggests that sexual dimorphism affects ancestry estimation, the most commonly used metric method to estimate ancestry—FORDISC— does not require the user to select a sex estimate prior to analysis. The FORDISC User's Guide (Jantz & Ousley, 2005, 2010) recommends that, after entering all cranial and postcranial measurements and ensuring they are correct and sample sizes are appropriate, all variables should be analyzed without any assumptions against all reference groups (all ancestry groups of both sexes). The most dissimilar groups (regardless of sex or ancestry) should then be removed based on typicality probabilities until there are only two to five groups remaining. Following these instructions, a prior estimation of sex has no impact on the analysis or results. However, in a more recent publication, Ousley and Jantz (2012) suggest that if additional information regarding an unknown individual is available, such as sex, than the analysis should be performed using either the all-male or all-female function, as including fewer groups increases accuracy. If the sex estimate is correct, this should facilitate a more accurate analysis; however, how incorrect sex estimation may affect other analyses in FORDISC is examined in more detail later in this chapter.

Care should be taken when choosing to compare an unknown individual to groups of only one sex. FORDISC will always classify an individual. For example, posterior probabilities may indicate that an individual has a 99% probability of being from a certain group as compared to the other groups. Low typicalities, on the other hand, can indicate that the actual group the individual is from is not represented in the reference samples, even if that group is a different sex but not necessarily a different ancestry.

While FORDISC remains most popular for metric sex estimation (Klales, 2013), analyses may often result in over-classification of females, under-classification of males, and extreme sex bias. Using combined ancestral groups with differing levels of sexual dimorphism can have significant implications for accurate sex estimation.

Morphological ancestry estimation methods based on macromorphoscopic traits are also popular and commonly accepted among anthropologists. These methods rely on traits described to be associated with specific populations, although it should be noted that no trait is found exclusively in any single population (Hefner, 2009; Hefner & Ousley, 2014). Statistical methods exist to assess the interaction of sex and ancestry on these categorical traits, such as ordinal regression analysis and analysis of covariance (Hefner, 2009; Hefner & Ousley, 2014). In this way, traits that are related specifically to ancestry may, in some cases, be separated from those related to sex. However, Klales and Kenyhercz (2015) found that there is a variation between sexes within one ancestry group in trait score frequencies, indicating that there may be some degree of sexual dimorphism present in specific traits that are also informative for ancestry. To date, no morphological method to estimate ancestry requires a known, or estimated, sex, making it difficult to assess how a misclassification of sex may impact ancestry estimation using morphological methods.

Sex and stature estimation

All aspects of the biological profile, including sex, age, and ancestry, are known to have some influence on stature. At the population level, male stature tends to be greater than female stature, although levels of sexual dimorphism, including height and skeletal robustness, vary among populations (e.g., the average stature for males and females, and the differences between these averages, vary significantly among groups). Stature can be estimated using one of two general approaches: (1) anatomical or full-skeleton methods where all skeletal elements that directly contribute to living height (cranium, vertebrae, sacrum, femur, tibia, talus, and calcaneus) are summed along with a soft tissue factor (e.g., Fully, 1956; Raxter, Auerbach, & Ruff, 2006) and (2) regression or mathematical methods where the relationships between a single bone, bone fragment, or multiple bones and height are used to estimate stature (e.g., Jantz & Ousley, 2005; Wilson, Nicholas, & Jantz, 2010). Due to the relatively direct relationship between size of the skeleton and height of an individual, anatomical methods have the advantage of being applicable to both sexes and any population.

Despite their general applicability and potential for high accuracy, anatomical methods are rarely used in forensic settings because of the necessity of having most, or all, of the necessary skeletal elements present and relatively free from modification (e.g., fracture, animal scavenging). In contrast, regression methods can be used in almost all cases, albeit with varying degrees of accuracy and precision. However, because the relationship between length of body segments and total height varies between the sexes and among populations, sex- and ancestry-specific equations are recommended.

Historically, population variation has been recognized in anthropological stature methods in the form of ancestry-specific equations. As early as 1929, Stevenson cautioned against using equations derived from northern Chinese samples with Rollet's, 1888 work on French cadavers (Rollet, 1888; Stevenson, 1929). Hrdlička (1939) and Trotter and Gleser (1952, 1958) also warned against combining formula from different populations. It can be surmised that some of these observed differences, and related advice to use stature equations specific to ancestry groups, may come from variations in sexual dimorphism between populations. Metric stature methods have been continuously developed for an increasing number of populations around the world (e.g., Auerbach & Ruff, 2010; Choi, Chae, Chung, & Kang, 1997; Genovés, 1967; Meadows & Jantz, 1992; Raxter et al., 2008). Most methods currently in use have sex-specific formulae that are associated with lower error estimates (e.g., Jantz & Ousley, 2005; Wilson et al., 2010).

In addition, FORDISC has become a popular method to estimate stature using postcranial measurements. Although a point estimate is generated, in forensic settings, estimated stature should always be presented as a range (90% or 95% prediction interval). The influence of incorrect sex assessment on stature prediction in FORDISC will be explored further using the sample of identified forensic cases evaluated later in this chapter.

Positively identified forensic cases

To assess the impact of misclassification of sex on other parameters of the biological profile, a review of nine forensic anthropological cases from the Mercyhurst University Forensic Case Databank was conducted. These select cases included positively identified individuals analyzed between 2010 and 2017. The biological profile data of each case, originally collected by the members of the Mercyhurst Forensic Anthropology Laboratory, were used to estimate ancestry, stature, and age for each individual. Craniometric data from seven cases (two were not available) were entered into FORDISC and analyses were conducted using the male groups, female groups, and all groups. Stature and age were calculated based on the assumption that sex was correctly estimated and then recalculated using the opposite (misclassified sex). As each forensic case is unique and may have had variable skeletal elements subjected to a myriad of taphonomic processes, not all aspects of the biological profile could be assessed for each individual.

Ancestry

A sample of seven of the positively identified cases from the Mercyhurst Forensic Anthropology Case Database (males 4, females 3) (Table 2) was metrically assessed for ancestry

Table 2 Estimated sex and ancestry of select positively identified individuals from the Mercyhurst University Forensic Case Databank using FORDISC 3.1.[a]

	All groups			Correct sex groups			Incorrect sex groups		
Positive identification	**Anc./sex est**	**PP[a]**	**F-typ**	**Anc./sex est**	**PP**	**F-typ**	**Anc./sex est**	**PP**	**F-typ**
White male	WM	0.99	0.96	WM	1.00	0.95	WF	1.00	0.25
White male	WM	1.00	0.27	WM	1.00	0.56	WF	1.00	0.02
White male	WM	0.65	0.63	WM	0.98	0.66	WF	0.99	0.12
White male	WM	0.99	0.15	WM	1.00	0.42	WF	0.99	0.00
White female	WF	0.61	0.13	WF	0.42	0.15	GTM	0.77	0.22
White female	WF	1.00	0.18	WF	1.00	0.18	GTM	0.58	0.04
White female	WF	0.61	0.41	WF	0.58	0.60	GTM	0.43	0.17

[a]Each analysis includes a different number of groups in the comparison based on the procedure outlined in the text (PP, posterior probability; F-typ, F-typicality).

using FORDISC 3.1 (Jantz & Ousley, 2005). All cases examined were of documented European ancestry. The craniometric data were used to examine the impact of incorrect sex selection on ancestry estimation. Once the data were entered, three sets of analyses were run: one using combined sexes (all groups), one using only female groups, and one using only male groups. For each analysis, measurements that were too high or too low (±3 standard deviations) were checked and, if deemed necessary, removed. Tests were run until no typicalities less than 0.05 remained or only two groups remained in the analysis.

In each analysis, including only groups of the incorrect sex resulted in measurements flagged as either too large or too small, particularly in areas often associated with sexual dimorphism, such as mastoid height. In all cases, when a male or female was analyzed using only groups of the incorrect sex, oddities were detected in the results; however, the pattern of atypical results was different between the sexes.

Positively identified males examined using all-female groups resulted in a classification of white females with high posterior probabilities, but low typicalities. In this scenario, it would be easy for a practitioner inexperienced with FORDISC to interpret the extremely high posterior probability—incorrectly—as a strong indication of correct group classification (i.e., the individual was highly likely to belong to that group). In each of the four cases analyzed, white males were classified as closest to white females in the set of reference groups used; however, the low typicalities for all groups indicate that the individual is not classifying well and is likely not represented by any of the groups included in the analysis. Based on this result, albeit a small sample, if there are high posterior probabilities associated with very low typicalities, it is important to consider the possibility of incorrect sex estimation. When all ancestry groups of both sexes were evaluated for the same individuals and following the same procedure outlined previously, the analyses all came down to a two-way comparison between white males and females. An all-group, combined-sex analysis and male-only analysis both resulted in high posterior probabilities, but the correct sex group comparison produced, as expected, higher typicalities.

When positively identified females are examined following the same procedure, the results are somewhat different. Compared to the male-only groups in FORDISC, these individuals classified most closely to the Guatemalan male group, with moderate to high posterior probabilities, but low typicalities. This suggests that incorrect sex assessment may have more significant repercussions for females mistaken for males than the other way around; however, a much larger sample, including those from other ancestry groups, should be examined before conclusions can be drawn. As with the positively identified males, a combined-sex analysis and an analysis using only females (the correct sex) produced generally high posterior probabilities, with the correct sex analysis associated with higher typicalities than the combined-sex analysis. In two of the three cases of positively identified females examined, the posterior probability decreased between the combined-sex group analysis and the correct sex analysis. This likely has to do with the number of groups included in the comparison.

This small survey of forensic cases hints that, in cases with unidentified individuals, the difference between using combined sex groups in FORDISC and groups of only one sex is relatively small, if the presumed sex estimate is accurate. Typicalities are also likely to be higher when groups of the correct sex are used. In contrast, if the incorrect sex is selected (in our sample particularly when females were mistaken for males), ancestry estimation of an individual may be affected. This is consistent with the known variation in sexual dimorphism between ancestries and the degree of variations present between males and females. These data indicate that when comparing an unknown individual to groups of only one sex, care need to be taken to interpret the posterior probabilities correctly and that close attention and critical thought should be paid to low typicalities.

Age

Age was estimated for all nine identified cases using TA (Boldsen et al., 2002). TA was used because it is the only method commonly employed by forensic practitioners that produces individualized age estimates. The estimates are based on the traits available for analysis and the collective suit of character states throughout the skeleton, combined with user-selected information about sex and ancestry. Because of this interplay of factors, the effect of a misclassification of sex will not be exactly the same for each individual, or category of individuals.

Age estimation results are presented in Table 3. The ages reported are 95% confidence intervals and maximum likelihood estimates from ADBOU (ver. 2.1.046) that were

Table 3 Estimated age (in years) of select positively identified individuals from the Mercyhurst University Forensic Case Databank.

Positive ID		Using correct sex		Using misestimated sex	
		Age interval	**Point estimate**	**Age interval**	**Point estimate**
Males	White, 19 years	17.9–21.8	17.9	18.5–22.2	18.5
	White, 19 years	16.7–23.9	20.2	17.0–24.2	20.5
	White, 29 years	24.5–34.5	29.1	23.8–37.6	29.5
	White, 32 years	28.7–47.9	36.3	26.6–47.3	34.8
	White, 49 years	31.1–57.1	41.0	30.0–57.2	41.5
Females	White, 17 years	15.3–23.5	19.6	15.7–23.8	19.8
	White, 29 years	21.7–33.5	26.7	21.9–33.2	26.9
	White, 30 years	16.3–25.8	20.7	16.1–25.7	20.6
	Black, 79 years	51.0–87.1	68.5	50.9–89.9	70.2

Ages reported are 95% confidence intervals and maximum likelihood estimates from ADBOU (ver. 2.1.046) using appropriate sex and ancestry reference samples and the forensic prior as all selected cases are documented or suspected homicides.

calculated using appropriate sex and ancestry reference samples and the forensic prior, as all included cases are known or suspected homicides. Maximum likelihood estimates produced using different sex classifications differed by less than 2 years in all cases, and 95% confidence intervals were similarly consistent. Thus, in these cases, a misestimation of sex would have little impact on the age estimation produced.

Based on the many nonstandard ways in which practitioners use techniques and combine data from multiple methods (Garvin & Passalacqua, 2012), it is impossible to predict the exact effect that misclassification will have; however, because the sex-related age estimation errors in the majority of commonly used techniques are theoretically small, the combined error is likely to be insignificant in a practical sense. Additional investigation in this area is needed.

Stature

Stature estimation results are presented in Table 4 for all nine individuals. The statures shown were calculated using FORDISC 3.1 (Jantz & Ousley, 2010) and are 90% prediction intervals for the equation with the highest R^2 value. In this small sample of positively identified cases, it is apparent that using regression equations for the incorrect, or misestimated, sex does impact the prediction interval and associated stature range that would be presented to law enforcement. For males, estimating stature with equations from a female reference sample lowered the height estimate by approximately 1 inch at the lower end of the range and approximately 0–2 inches on the upper end. The change in stature point estimate ranged from 0.55 to 2.45 inches, with an average decrease in stature of 1.2 inches For European females (three of the four individuals

Table 4 Estimated stature of select positively identified individuals from the Mercyhurst University Forensic Case Databank.

Positive ID		Using correct sex Stature range		Using misestimated sex Stature range	
Males	White, 19 years	66.3–72.9 in.	(5′6″–6′1″)	64.9–71.1 in.	(5′5″–5′11″)
	White, 19 years	66.8–73.3 in.	(5′7″–6′1″)	65.7–71.9 in.	(5′6″–6′0″)
	White, 29 years	67.7–74.2 in.	(5′8″–6′2″)	67.2–73.6 in.	(5′7″–6′2″)
	White, 32 years	65.5–72.7 in.	(5′6″–6′1″)	65.1–72.0 in.	(5′5″–6′0″)
	White, 49 years	67.7–74.2 in.	(5′8″–6′2″)	66.5–72.7 in.	(5′7″–6′1″)
Females	White, 17 years	56.4–63.2 in.	(4′8″–5′0″)	58.2–64.9 in.	(4′10″–5′5″)
	White, 29 years	53.8–60.2 in.	(4′6″–5′0″)	56.5–63.2 in.	(4′9″–5′3″)
	White, 30 years	60.3–66.4 in.	(5′0″–5′6″)	61.8–68.3 in.	(5′2″–5′8″)
	Black, 79 years	67.2–72.2 in.	(5′7″–6′0″)	63.1–69.3 in.	(5′3″–5′9″)

Stature is the 90% prediction interval reported by FORDISC (ver. 3.1).

evaluated), the change in stature when equations based on male reference samples were used was more variable. The use of a male reference group increased stature ranges by 2–3 inches on the lower end and 2–5 inches on the upper end.

The fourth female evaluated, whose documented ancestry was listed as black, highlights a potentially important point from this assessment. The other individuals in this small sample are of European ancestry. Given what is known about the variation in sexual dimorphism among populations and differences in the relationship between body proportions and total height between males and females, it is likely that these results would be slightly different using samples from other ancestry groups. The black female in this group shows an unexpected decrease in average stature when applying male standards. This is potentially the result of a difference in the relationship between body proportions and height in this population as compared to the European group, but should be systematically investigated using a large documented sample.

These data highlight the strong possibility that the error introduced in estimated stature by an incorrect sex estimation method is not the same for males and females. Additionally, variation among populations may significantly shift or even reverse the patterns seen here when similar tests are conducted with larger and more samples. Although it should be noted that the use of combined sex and ancestry equations in FORDISC using the "any" option will increase the width of the prediction interval produced because of the increased variation present in the reference sample (Jantz & Ousley, 2010). This means that stature estimates, which are often already wide and not practically useful for reducing the pool of potential matches in medicolegal contexts, will be even less effective.

Although the estimates in this sample are only different by a few inches, in most cases, the stature range would be changed enough that individuals outside of that range might not be considered when attempting to identify an unknown individual. More importantly, however, it could be argued that the far more important issue in terms of identification would be the incorrect sex estimate.

Case study

The select cases presented from the Mercyhurst Forensic Case Databank provided examples of the effect of misclassification of sex on particular components of the biological profile (age and stature). This case provides an example where the results of sex estimation were ambiguous and the potential misclassification of sex may have important implications on the identification of this individual who currently remains unidentified.

An individual walking along a riverbank in Pennsylvania discovered what appeared to be partially clothed, partially skeletonized human remains. The individual located several scattered skeletal elements and stacked them near the body before contacting the state police. The Mercyhurst Forensic Scene Recovery Team was later contacted to conduct a forensic archaeological recovery and subsequent anthropological analysis to assess the unknown individual's biological profile and to assess the remains for skeletal trauma.

Once recovered and inventoried, the remains were found to represent a single incomplete adult. Elements not recovered at the scene included the cranium, mandible, right radius, and the complete upper left limb. Morphological and metric analyses were used to estimate sex. Morphological assessment of the innominates revealed asymmetric expression of typical sex characteristics. The pubic bone was relatively short, with a small pubic-to-ischium-length ratio (Bruzek, 2002) (Fig. 2A). The ilia were vertically positioned; obturator foramina were triangular (Fig. 1A); and the acetabula were relatively large (Rogers & Saunders, 1994). Although asymmetrical, the shape of greater sciatic notches was characteristically male as illustrated by the posterior chord being shorter than the anterior chord, and overall the angle was not wide (Bruzek, 2002; Rogers & Saunders, 1994) (Fig. 2C). There was also an asymmetrical expression of the subpubic contour with a greater concavity on the left side than on the right side (Fig. 2A). The slight subpubic concavity (although asymmetrical), along with a slight ventral arc, is present on both innominates and is more characteristic of female individuals (Phenice, 1969) (Fig. 2A).

Using scores from the medial aspect of the ischio-pubic ramus, subpubic contour, and ventral arc scores from the Klales, Ousley, and Vollner (2012) logistic regression functions provides a 75.6% probability for the left innominate and a 51.5% probability for the right innominate that the individual is female. It should be noted that these results are only slightly better than chance and are not a strong indication that this individual is female. Overall, when the morphology of the pubis, ischium, greater sciatic notch, obturator foramen, iliac blades, and acetabula are considered in conjunction with inconclusive results of the Klales et al. (2012) method, the morphological characteristics of innominates are more consistent with a male individual.

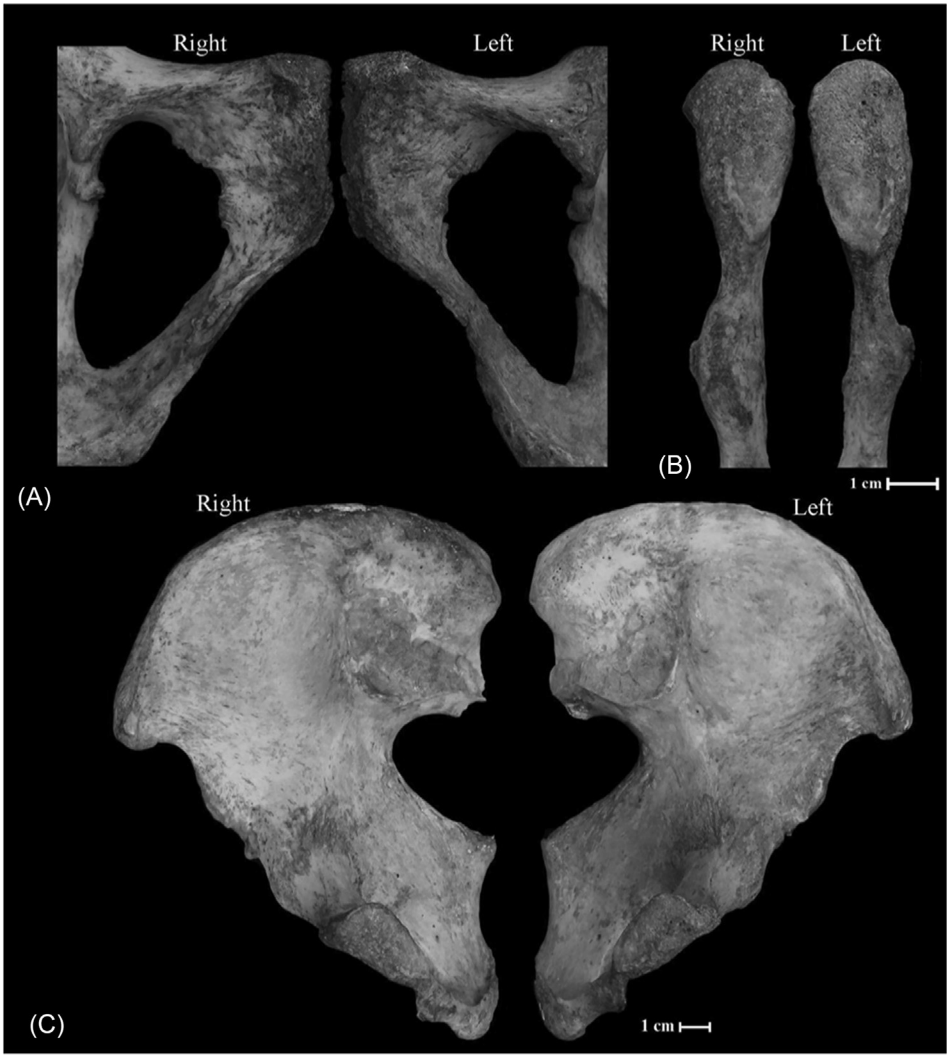


Fig. 2 Right and left innominates from the Mercyhurst University forensic case study. (A) Ventral surface of the left and right innominates showing the ventral arc area, asymmetric subpubic contour, triangular obturator foramina, and relatively short pubic bones. (B) Medial aspect of the ischio-pubic ramus of the left and right innominates. (C) Medial views of the left and right innominates showing greater sciatic notch morphology and overall asymmetry. *(Photograph courtesy of the Mercyhurst Forensic Anthropology Laboratory Director, Dennis C. Dirkmaat.)*

Thirty-six standard postcranial measurements were also recorded for this individual. These were compared to 735 individuals of known ancestry and sex, through LDFA in FORDISC 3.1 (Jantz & Ousley, 2010), following the guidelines outlined in Ousley and Jantz (2012). All forensic reference samples were entered into the analysis: US black and white individuals of both sexes. Wilk's lambda forward-stepwise selection (TYP = 0.05; W-step = 0.005) was utilized to prevent sample overfitting (Ousley & Jantz, 2012). The sample groups with F-typicalities lower than the 0.05 threshold were removed to narrow down the comparison to the most similar groups. The results of this analysis indicate that the postcranial skeleton was most similar to the US black male group with a posterior probability of 96.1% (F-typicality = 0.991) compared to the white male group with a posterior probability of 3.9% (F-typicality = 0.497) when using nine measurements. When only two ancestral groups were analyzed, this individual was 25 times more likely to be a US black male than a US white male.

Sex was also estimated using the Spradley and Jantz (2011) discriminate function for the humerus, which had the highest classification rate for American black males. This analysis suggested the individual was male with a 93.84% accurate classification rate. The combination of morphological traits and results of metric analysis indicates that this individual was likely a male. In addition, postcranial metrics were most consistent with US black males. However, given the ambiguity of several pelvis features and lack of a skull to contribute to sex estimation, it was possible that sex for this individual could have been misestimated. Had that been the case, other parameters of the biological profile may have been impacted.

For example, based on the sex and ancestry results, stature of the decedent was estimated using regression equations provided for the 20th-century forensic black male reference group in FORDISC 3.1 (Jantz & Ousley, 2010). The regression equation using the maximum fibula length, maximum humerus length, and maximum tibia length displayed the highest correlation with stature and provided an estimate between 5′5″ and 6″. However, had the results been run using regression equations provided for the 20th-century forensic black female reference group, the regression equation, using the maximum clavicle length, innominate height, and maximum tibia length, would have displayed the highest correlation with stature and suggested an estimate of between 5′9″ and 6′2″. Aside from the major issue of incorrect sex in a forensic context, this could mean the difference between either inclusion or exclusion of individuals based on their height. In this case, the difference is significant, and in practical terms, the search would shift from looking for a male of average height to a tall female. It must be kept in mind that the individual represented in this case study is unidentified. In such a case, if a substantial concern for sex estimation is raised, it may be prudent to use the "any" setting in the FORDISC stature section. In this particular case, the formulae generated from the "any" setting provide a stature range of 5′6″ to 6′1″.

When the prior sex and ancestry (black male) assessments were used in the TA software (ADBOU, ver. 2.1.046) to estimate age, an estimate of 24.6–40.1 years with a maximum likelihood of 31.0 years was produced. Keeping ancestry the same, but changing the sex classification to female produced an estimate of 23.5–38.4 years with a maximum likelihood of 29.5 years. Although present, the differences in both point and interval estimates are minimal. In this case, as with stature, the most conservative approach would be to assume that both sex and ancestry are unknown. Using the "unknown" selections for both sex and ancestry combines individuals from each of the reference samples to estimate age and produces an estimate of 23.4–40.3 years with a maximum likelihood of 30.2 years, which encompasses the intervals provided for both male and female estimates.

In the original case report, not written by either of this chapter's authors, age was estimated using multiple methods, including those for pubic symphyses, iliac auricular surfaces, and sternal rib ends. The age ranges from each method were charted, and wide and narrow ranges were generated based on the areas of overlap and practitioner experience.

Although the cranium typically contributes little to age estimation, its absence in this case complicated age estimation because ancestry had to be assessed from the postcranial skeleton. Ancestry and sex are not considered in either of the methods for the iliac auricular surface (Buckberry & Chamberlain, 2002; Osborne et al., 2004). For the sternal rib ends, alternative age intervals (mean ± 2SD) are available for black and white males and females (İşcan et al., 1987). In this case, assuming that the black ancestry assessment is correct, a misestimation of sex would result in a significant difference in the estimated age. For Phase 3, black males have a mean of 24.9 years with an approximate 95% confidence interval of 18.42–31.38 years, whereas black females have only a point estimate age of 21 years because it is based on a sample size of one individual. Although a misestimation of sex affects the age, a much larger error would be introduced by an inaccurate ancestry estimate. Not only are the age intervals produced for the black and white groups different, the magnitude of error produced by an inaccurate sex estimation within each group is also different. An incorrect ancestry assessment would also influence the age intervals produced from the pubic symphysis for males and females because, while alternative intervals are available for black males, comparable intervals for black females are not, so the same methods could not be used.

Ultimately, in this case, the effect of an inaccurate sex estimate would be dramatically reduced by the multimethod, chart-based approach interpreted using practitioner experience. However, if a practitioner chose to use only a single method or combine the results in a different way, the effect of misestimation of ancestry or sex could have a more dramatic effect.

Conclusion

Sex unquestionably impacts multiple parameters of the biological profile; however, the magnitude of errors introduced by a misclassification of sex is largely defined by the methods used and the way in which they are combined to generate a final assessment. It is also imperative that the selected methods be applied correctly, although, as with the case of age estimation, "correct" method use is not always clear.

In our assessment of a small sample of positively identified forensic cases, the greatest effects of misclassification of sex were seen in the estimation of stature where sex-specific equations are highly recommended to obtain the most useful results. Using an incorrect sex-specific equation seems to result in a relatively small, but practically significant, difference of one to several inches in the estimated stature range. Although a few inches may seem insubstantial, this discrepancy could mean the difference between inclusion or exclusion of individuals in a forensic context where the identity of the person is unknown. Thus, we recommend that in cases where sex may be ambiguous, stature ranges be expanded to include the full estimated range for both sexes.

For the same sample of select cases, the effect of misclassification of sex on age estimates produced using TA (Boldsen et al., 2002) was negligible. Selecting "unknown" sex in the software will produce a conservative estimate based on a combined sample of males and females, so this selection should be used in cases when sex is ambiguous. For combinations of other traditional methods when sex cannot be confidently estimated, it is recommended that wider intervals spanning possible estimates for males and females be provided.

Morphological ancestry estimation methods currently are not sex-specific; however, differences between sexes in traits associated with ancestry have been noted (Klales & Kenyhercz, 2015). Incorrect sex assessment can have significant repercussions using FORDISC, and potentially alter ancestry estimation, particularly if a female is mistaken for a male. There are, however, distinct clues (high posterior probabilities, low typicalities) that may indicate something is wrong with the FORDISC analysis. These features generally indicate that the individual is not represented in the select reference samples, which can be an issue with sex and/or ancestry. If there is any doubt regarding the sex of an individual, an analysis in FORDISC examining combined sex groups often yields similar posterior probabilities, but with slightly higher typicalities, to an analysis using only correct sex groups.

In summary, this investigation supports the assertion that sex estimation should typically be the first step in the construction of a biological profile. Error introduced by misclassification of sex will vary based on the methods used, how they are combined into a final estimate, and the actual age, sex, and ancestry of the person being evaluated, but the overall impact of misestimation is likely to be minimal. This is somewhat contradictory to the commonly held belief that sex is the most critical of all the biological profile parameters because of its significant influence on other areas. Arguably, in many contexts, the most severe problem caused by misestimation of sex in the construction of a biological profile is likely to be the incorrect sex assessment itself.

References

Auerbach, B. M., & Ruff, C. B. (2010). Stature estimation formulae for indigenous North American populations. *American Journal of Physical Anthropology*, *141*(2), 190–207.

Boldsen, J. L., Milner, G. R., Konigsberg, L. W., & Wood, J. W. (2002). Transition analysis: A new method for estimating age from skeletons. In R. Hoppa & J. Vaupel (Eds.), *Paleodemography: Age distributions from skeletal samples* (pp. 73–106). Cambridge: Cambridge University Press.

Brooks, S. T., & Suchey, J. M. (1990). Skeletal age determination based on the os pubis: A comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Human Evolution*, *5*(3), 227–238.

Bruzek, J. (2002). A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology*, *117*(2), 157–168.

Buckberry, J. L., & Chamberlain, A. T. (2002). Age estimation from the auricular surface of the ilium: A revised method. *Journal of Physiological Anthropology*, *119*(3), 231–239.

Cabo, L. L., Brewster, C. P., & Azpiazu, J. L. (2012). Sexual dimorphism: Interpreting sex markers. In D. C. Dirkmaat (Ed.), *A companion to forensic anthropology* (pp. 248–286). Malden, MA: Wiley and Blackwell [Chapter 13].

Choi, B. Y., Chae, Y. M., Chung, I. H., & Kang, H. S. (1997). Correlation between the postmortem stature and the dried limb-bone lengths of Korean adult males. *Yonsei Medical Journal*, *38*, 79–85.

Christensen, A. M., Passalacqua, N. V., & Bartelink, E. J. (Eds.), (2014). *Forensic anthropology: Current methods and practice*. New York: Academic Press.

Cox, M. (2000). Ageing adults from the skeleton. In M. Cox & S. Mays (Eds.), *Human osteology in archaeology and forensic science* (pp. 61–82). Cambridge, UK: Cambridge University Press.

Dirkmaat, D. C., Cabo, L. L., Ousley, S. D., & Symes, S. A. (2008). New perspectives in forensic anthropology. *American Journal of Physical Anthropology*, *51*(S47), 33–52.

Dwight, T. (1890). The closure of the cranial sutures as a sign of age. *Boston Medical and Surgical Journal*, *122* (17), 389–392.

Falys, C. G., & Lewis, M. E. (2011). Proposing a way forward: A review of standardisation in the use of age categories and ageing techniques in osteological analysis (2004–2009). *International Journal of Osteoarchaeology*, *21*(6), 704–716. https://doi.org/10.1002/oa.1179.

Fully, G. (1956). New method of determination of height. *Annales de Me´dine Le´gale, Criminologie, Police Scientifiqueet Toxicologie*, *36*(5), 266–273.

Garvin, H. M., & Passalacqua, N. V. (2012). Current practices by forensic anthropologists in adult skeletal age estimation. *Journal of Forensic Sciences*, *57*(2), 427–433. https://doi.org/10.1111/j.1556-4029.2011.01979.x.

Genovés, S. (1967). Proportionality of the long bones and their relation to stature among Mesoamericans. *American Journal of Physical Anthropology*, *26*(1), 67–78.

Gilbert, B. M., & McKern, T. W. (1973). A method for aging the female os pubis. *American Journal of Physical Anthropology*, *38*(1), 31–38.

Guyomarc'h, P., & Bruzek, J. (2011). Accuracy and reliability in sex determination from skulls: a comparison of Fordisc 3.0 and the discriminant function analysis. *Forensic Science International*, *208*(1–3). https://doi.org/10.1016/j.forsciint.2011.03.011. 180.e1–180.e6.

Hartnett, K. M. (2010). Analysis of age-at-death estimation using data from a new, modern autopsy sample. Part II. Sternal end of the fourth rib. *Journal of Forensic Sciences*, *55*(5), 1152–1156.

Hefner, J. T. (2009). Cranial nonmetric variation and estimating ancestry. *Journal of Forensic Sciences*, *54*(5), 985–995.

Hefner, J. T., & Ousley, S. D. (2014). Statistical classification methods for estimating ancestry using morphoscopic traits. *Journal of Forensic Sciences*, *59*(4), 883–890.

Hershkovitz, I., Latimer, B., Dutour, O., Jellema, L. M., Wish-Baratz, S., Rothschild, C., et al. (1997). Why do we fail in aging the skull from the sagittal suture? *American Journal of Physical Anthropology*, *103*(3), 393–399. https://doi.org/10.1002/(SICI)1096-8644(199707)103:3<393::AID-AJPA8>3.0.CO;2-R.

Hrdlička, A. (1939). *Practical anthropometry*. Philadelphia, PA: Wistar.

Igarashi, Y., Uesu, K., Wakebe, T., & Kanazawa, E. (2005). New method for estimation of adult skeletal age at death from the morphology of the auricular surface of the ilium. *American Journal of Physical Anthropology*, *128*(2), 324–339.

İşcan, M. Y., & Loth, S. R. (1986a). Determination of age from the sternal rib in white males: A test of the phase method. *Journal of Forensic Sciences*, *31*(1), 122–132.

İşcan, M. Y., & Loth, S. R. (1986b). Determination of age from the sternal rib in white females: A test of the phase method. *Journal of Forensic Sciences*, *31*(3), 990–999.

İşcan, M. Y., Loth, S. R., & Wright, R. K. (1984a). Metamorphosis at the sternal rib end: A new method to estimate age-at-death in white males. *American Journal of Physical Anthropology*, *65*(2), 147–156.

İşcan, M. Y., Loth, S. R., & Wright, R. K. (1984b). Age estimation from the rib by phase analysis: White males. *Journal of Forensic Sciences*, *29*(4), 1094–1104.

İşcan, M. Y., Loth, S. R., & Wright, R. K. (1985). Age estimation from the rib by phase analysis: White females. *Journal of Forensic Sciences*, *30*(3), 853–863.

İşcan, M. Y., Loth, S. R., & Wright, R. K. (1987). Racial variation in the sternal extremity of the rib and its effect on age determination. *Journal of Forensic Sciences*, *32*(2), 452–466.

Jantz, R. L., & Ousley, S. (2005). *FORDISC 3: Computerized forensic discriminant functions*. Knoxville, TN: The University of Tennessee.

Jantz, R. L., & Ousley, S. D. (2010). *FORDISC 3.1: Computerized forensic discriminant functions. Version 3.1*. Knoxville, TN: University of Tennessee.

Katz, D., & Suchey, J. M. (1986). Age determination of the male os pubis. *Journal of Physiological Anthropology*, *69*(4), 427–435.

Katz, D., & Suchey, J. M. (1989). Race differences in pubic symphyseal aging patterns in the male. *Journal of Physiological Anthropology*, *80*(2), 167–172.

Klales, A. R. (2013). Current practices in physical anthropology for sex estimation in unidentified, adult individuals. *American Journal of Physical Anthropology*, *150*(S56), 168.

Klales, A. R., & Kenyhercz, M. W. (2015). Morphological assessment of ancestry using cranial macromorphoscopics. *Journal of Forensic Sciences*, *60*(1), 13–20.

Klales, A. R., Ousley, S. D., & Vollner, J. M. (2012). A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *American Journal of Physical Anthropology*, *149*(1), 104–114.

Lovejoy, C. O., Meindl, R. S., Pryzbeck, T. R., & Mensforth, R. P. (1985). Chronological metamorphosis of the auricular surface of the ilium: A new method for the determination of adult skeletal age at death. *American Journal of Physical Anthropology*, *68*(1), 15–28.

McKern, T. W., & Stewart, T. D. (1957). *Skeletal age changes in young American males analysed from the standpoint of age identification*. Technical Report EP-45. Natick, 1v1A: Quartermaster Research and Development Command.

Meadows, L., & Jantz, R. L. (1992). Estimation of stature from metacarpal lengths. *Journal of Forest Science*, *37* (1), 147–154.

Meindl, R. S., & Lovejoy, C. O. (1985). Ectocranial suture closure: A revised method for the determination of skeletal age at death based on the lateral-anterior sutures. *American Journal of Physical Anthropology*, *68*(1), 57–66.

Milner, G. R., & Boldsen, J. L. (2012). Transition analysis: A validation study with known-age modern American skeletons. *American Journal of Physical Anthropology*, *148*(1), 98–110.

Nawrocki, S. P. (1998). Regression formulae for estimating age at death from cranial suture closure. In K. J. Reichs (Ed.), *Forensic osteology: Advances in the identification of human remains* (pp. 276–292). Springfield: Charles C. Thomas.

Osborne, D. L., Simmons, T. L., & Nawrocki, S. P. (2004). Reconsidering the auricular surface as an indicator of age at death. *Journal of Forensic Sciences*, *49*(5), 1–7.

Ousley, S. D., & Jantz, R. L. (2012). Fordisc 3 and statistical methods for estimating sex and ancestry. In D. C. Dirkmaat (Ed.), *A companion to forensic anthropology* (pp. 248–286). Malden, MA: Wiley and Blackwell.

Phenice, T. (1969). A newly developed visual method for sexing the os pubis. *American Journal of Physical Anthropology*, *30*(2), 297–302.

Raxter, M. H., Auerbach, B. M., & Ruff, C. B. (2006). Revision of the fully technique for estimating statures. *American Journal of Physical Anthropology*, *130*(3), 374–384.

Raxter, M. H., Ruff, C. B., Azab, A., Erfan, M., Soliman, M., & El-Sawaf, A. (2008). Stature estimation in ancient Egyptians: A new technique based on anatomical reconstruction of stature. *American Journal of Physical Anthropology*, *136*(2), 147–155.

Rogers, T., & Saunders, S. (1994). Accuracy of sex determination using morphological traits of the human pelvis. *Journal of Forensic Sciences*, *39*(4), 1047–1056.

Rollet, E. (1888). *De la mesuration des os longs des membres. Thèse pour le dotorat en médicine, 1st series. Vol. 43*. (pp. 1–128).

Spradley, M. K., & Jantz, R. L. (2011). Sex estimation in forensic anthropology: Skull versus postcranial elements. *Journal of Forensic Sciences*, *56*(2), 289–296.

Spradley, M. K., Jantz, R. L., Robinson, A., & Peccerelli, F. (2008). Demographic change and forensic identification: Problems in metric identification of Hispanic skeletons. *Journal of Forensic Sciences*, *53*(1), 21–28.

Stevenson, P. H. (1929). On radial differences in stature long bone regression formulae, with special reference to stature reconstruction formulae for the Chinese. *Biometrika*, *21*, 303–321.

Suchey, J. M., Brooks, S. T., & Katz, D. (1988). *Instructions for use of the Suchey-Brooks system for age determination of the female os pubis. Instructional materials accompanying female pubic symphysial models of the Suchey-Brooks system*. CO, France Casting: Fort Collins.

Todd, T. W. (1920). Age changes in the pubic bone. I. The male white pubis. *American Journal of Physical Anthropology*, *3*(3), 285–334.

Todd, T. W. (1921). Age changes in the pubic bone. *American Journal of Physical Anthropology*, *4*(1), 1–70.

Trotter, M., & Gleser, G. C. (1952). Estimation of stature from long bones of American Whites and Negroes. *American Journal of Physical Anthropology*, *10*(4), 463–514.

Trotter, M., & Gleser, G. C. (1958). A re-evaluation of estimation of stature taken during life and of long bones after death. *American Journal of Physical Anthropology*, *16*(1), 79–123.

Wilson, R. J., Nicholas, M. A., & Jantz, L. M. (2010). Evaluation of stature estimation from the database for forensic anthropology. *American Journal of Physical Anthropology*, *55*(3), 684–689.

CHAPTER 6

Sex estimation using pelvis morphology

Alexandra R. Klales
Forensic Anthropology Program, Department of Sociology and Anthropology, Washburn University, Topeka, KS, United States

Introduction

The articulated pelvis consists of the left and right innominates, the sacrum, the coccyx, and three major articulations: sacroiliac (left and right) and the pubic symphysis. It functions to support and protect the internal viscera, transmit weight between the upper body and lower body, and form the major joint for the articulation of the lower limbs. Both the innominates and the sacrum exhibit sexually dimorphic differences in size and shape that begin at the onset of puberty, when females show increased growth in the acetabular and pubic areas relative to the growth in other pelvic regions (Coleman, 1969). The differences primarily center on the trade-off between efficient bipedal locomotion, which requires a narrower pelvis, and the need for a wider pelvis in females to accommodate childbirth, commonly referred to the "obstetrical dilemma" (Washburn, 1960). A direct correlation has been found between neonatal brain size relative to the pelvic canal and the level of sexual dimorphism in a species (Ridley, 1995). Recent work by Fischer and Mitteroecker (2015) has called into question the obstetric dilemma due to the complex relationship between body size factors (head size and stature) and pelvic dimensions; however, the selective pressures of childbirth in females likely continue to contribute to the complex patterns of sexual dimorphism in male and female pelvic forms (see Ruff, 2017 for a more detailed discussion). Lastly, like many parts of the skeleton, the traits of the pelvis are highly heritable, but also demonstrate some phenotypic plasticity due to the environment (Fischer & Mitteroecker, 2015).

The pelvis has long been considered the best skeletal indicator of sex and has, therefore, been widely documented in the anatomical, anthropological, gynecological, and osteopathic literature. An extensive list of sexually dimorphic pelvis features is contained within these works; however, selection of which traits to use and when is largely dependent on the preferences and experiences of the observer. The Workshop of European Anthropologists (WEA) published standards for age and sex estimation in 1980. They recommended the use of 10 traits including the preauricular sulcus, greater sciatic notch, pubic angle, arc compose, obturator foramen, ischial body/tuberosity, ilium fossa, iliac crest, and the pelvis major and minor for morphological pelvic sex estimation (Table 1).

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00006-7

Table 1 Morphological pelvis traits used for sex estimation.

Acetabulum position	Pelvic inlet
Acetabulum size	Pelvis shape
Auricular surface height	Preauricular sulcus
Composite arch	Pubis shape
Dorsal pitting	Sacral segments
Ilium flaring	Sacrum shape
Ilium shape	Sciatic notch
Inferior pelvis	Subpubic angle
Ischiopubic proportion	Subpubic concavity/contour
Ischiopubic ramus	True pelvis shape
Length of sacroiliac joint	Ventral arc
Muscle markings	Visibility of sacroiliac joint
Obturator foramen shape	

In 1994, *Standards for Data Collection from Human Remains* (Buikstra & Ubelaker, 1994) suggest the use of Phenice's (1969) three traits, the ventral arc, subpubic concavity, and medial aspect of the ischiopubic ramus, along with the greater sciatic notch and preauricular sulcus for pelvic sex estimation (Table 1). That same year, Rogers and Saunders (1994) presented a comprehensive list of useful morphological pelvic traits and go one step further to test the accuracy and reliability of each of these traits (Table 1). The authors then proceeded to rank the 17 traits based on an overall effectiveness score that was a combined total of the accuracy and precision scores. One flaw in this approach is that no attempt was made to balance the validity and reliability scores, so some traits had high overall effectiveness rankings, but either the accuracy or precision scores were quite low. For example, sacrum shape scored fourth overall in effectiveness ranking; however, the validity rank (first) was considerably higher than the reliability rank (10th). The traits with the top 10 overall rank scores according to Rogers and Saunders (1994) are the ventral arc (first), obturator foramen, true pelvis, sacrum shape, subpubic concavity angle, pubis shape, muscle markings, dorsal pitting and acetabulum (tied for eighth place), preauricular sulcus, and the sacrum (posterior view) (tied for 10th place). Both the sciatic notch, suggested by Buikstra and Ubelaker (1994), and the ischiopubic ramus included in Phenice (1969) failed to make the top-10 list. Noticeably absent are also several of the traits included by the WEA (1980). Despite this ranking system, a survey of biological anthropologists (see Chapter 2 in this volume) indicated an overwhelming preference for the traits included in *Standards* and the three developed by Phenice (1969).

Below you will find descriptions of the top traits listed by Rogers and Saunders (1994), as well as other morphological pelvic traits commonly utilized in sex estimation. Recommendations for use or disuse are also included.

Morphological pelvis traits

Ventral arc

The utility of the ventral arc was recognized very early on by Cleland (1889) who identified that sex differences existed. The trait was formally introduced into a sexing method by Phenice (1969) along with the medial aspect of the ischiopubic ramus and the subpubic concavity (see more below on these two traits). He specified that the presence of a ventral arc indicated a female, while the absence indicated a male; however, his work lacks an explanation on why it is present in one sex versus the other. The ventral arc serves as the muscle attachment site for the gracilis and the adductor brevis and magnus muscles (Budinoff & Tague, 1990; Todd, 1921) and, therefore, can and should be found in both males and females (Anderson, 1990). Buikstra and Meilke (1985) and Naňko, Šedý, and Jarolím (2007) suggest that the arc is due to the crus penis and crus clitoris, and Bass (1987) suggests that it is the arcuate ligament attachment site. Debate exists as to whether these muscle attachment sites vary between the sexes (e.g., Budinoff & Tague, 1990) but, most generally agree that the orientation and position of the attachment site varies between the sexes likely due to the relatively wider pelvis in females (Anderson, 1990; Klales, Ousley, & Vollner, 2012; Sutherland & Suchey, 1991). In females, the arc is typically present at an angle of greater than 25 degrees in relation to the symphyseal face (Fig. 1, left image, black arrows) (Klales et al., 2012). In Phenice's (1969) original work, the subsequent modifications of his method by Klales et al. (2012), and in Rogers and Saunders (1994), the ventral arc had the highest overall classification accuracy (>85.0%) of the three traits. The low intra- and interobserver error associated with scoring this trait make it ideal for accurate sex estimation when available for analysis (Klales et al., 2012; Phenice, 1969; Rogers & Saunders, 1994).

Fig. 1 Left: Right innominate of a female individual showing the ventral arc presence and angle *(black arrows)*, subpubic concavity *(black line)*, and overall rectangular pubis shape. Right: Right innominate of a male showing a lack of a ventral arc, convex subpubic contour, and overall triangular pubis shape. *(Photos courtesy of Susie Athey.)*

Obturator foramen shape

The obturator foramen forms with the complete fusion of the acetabulum around puberty (between 11 and 17 years) (Baker, Dupras, & Tocheri, 2005). The male morphology is described as being larger and more ovoid with the greatest diameter obliquely (Gray, 1858). The female form is contrasted as smaller, more triangular, and taller (Gray, 1858). These shape differences arise from differential pubertal growth trajectories that result in a greater pubis length in females and a greater ischium length in males. While this has been perpetuated in the literature since Ackerman (1788) first described the sex differences, very few studies have actually tested the ability to visually distinguish between the morphologies (see Bierry, Le Minor, & Schmittbuhl, 2010 for a more detailed discussion of this history). Verneau (1875) very early on indicated that there is "no truth" in the purported shape differences between the sexes, and there is a significant overlap in transitional shapes; however, his work has largely been ignored (as cited in Bierry et al., 2010). One notable exception to the study of obturator foramen shape is Rennie (2018), who created a new five-point ordinal scoring system that ranged from oval (+2) to triangular (−2). Rennie (2018) found only moderate interobserver agreement in scoring and widely variable accuracy rates depending on the sample (58%–78%).

To better understand the visual shape differences in the obturator foramen, geometric morphometric (GMM) methods have been employed. Using elliptical Fourier analysis (EFA), Cline (2015) demonstrated shape (triangular vs. ovoid) differences between the sexes that could be used to accurately predict sex; however, the author cautions that shape alone should not be used due to low accuracy rates (<67.6%). Bierry et al. (2010) found similar results using EFA of just shape, but cautions that using a qualitative assessment of oval vs. triangular is insufficient. Assessing the trait visually, without the aid of these quantitative approaches, makes it likely more difficult to discern between the sexes, and differentiating transitional forms is likely not feasible. For the shape of the obturator foramen to be used as a morphological indicator of sex, without the aid of morphometric approaches, it needs to be more extensively validated for accuracy and reliability. Until this happens, use of this trait is not recommended.

True (lesser) pelvis

The lesser or true pelvis, also sometimes referred to as the pelvic cavity, refers to the area below or inferior to the pelvic inlet. Rogers and Saunders (1994) describe the male true pelvis as small, while the female expression is more often shallow and spacious. Assessment of this feature requires the ability to re-articulate the innominates and sacrum. While the overall rank of the pelvic inlet itself was low (15th) on Roger's and Saunder's (1994) list, a more detailed discussion is warranted because pelvis shape overall, shape of the greater vs. lesser pelvis, and inlet and outlet shape/size are all interconnected.

The pelvic inlet separates the pelvis into the greater/false pelvis (comprising the ala of the ilium above the arcuate line) and the lesser/true pelvis (Brown, 2010). The pelvic inlet is bounded by the sacral promontory, arcuate line, iliopubic eminence, pectineal line, pubic crest, and the pubic symphysis. It is widest in the medio-lateral direction. The pelvic outlet is widest in the anterio-posterior direction and is situated between the ischial tuberosities, sacral segment five, and the inferior edge of the pubic symphysis. Leong (2006) suggests that canalization of growth occurs in some pelvic dimensions (sacrum transverse diameter and ilium/ischium breadth) during puberty due to stabilizing selection, while other areas exhibit increased sexual dimorphism due to disruptive selection (inter-acetabular diameter, iliac spines, pubis length, and ilium height). These differences essentially create a larger pelvic inlet and outlet in females.

Turner (1885) described three categories of overall pelvic shape: dolichopellic/anthropoid, mesatipellic/gynecoid, and platypellic/platypelloid. To these classifications, Thoms (1946) added a fourth shape of brachypellic/android (see Delprete, 2017 and references within for more details on the history of pelvic shape classification). Females are typically described as gynecoid, while males are described as android. The gynecoid pelvis is described as wide, medio-laterally broad, with a short sacrum, and nonprominent ischial spines. The android pelvis is described as narrow, heart-shaped, with obstructions to the birth canal including prominent ischial spines and a long, anteriorly curved sacrum. However, Delprete (2017) found that the android or heart shape was pervasive in both sexes and was less sexually dimorphic than previous publications have suggested (cf. Abitbol, 1996; Burden & Simons, 2004; Drake, Vogl, & Mitchell, 2005).

In the false/greater pelvis, males are, on average, larger and more robust than females in the same population due to adaptations for supporting a larger body structure overall. The male pelvis is longer superior-inferiorly, sometimes referred to as higher or taller, than the female pelvis, which is, in turn, wider in the medio-lateral direction and shallower. Some research has suggested that both pelvic height and breadth are correlated with weight and/or stature, which would, therefore, impact false pelvis shape and the shape of the inlet (Moerman, 1981). Complicating the matters further are the impact of population differences on pelvis shape and the role of etiological factors (e.g., cultural, environmental, genetic) (Leong, 2006). At present, morphological assessment of the greater and lesser pelvis, including the inlet and outlet, should be limited due to these myriad factors. Metric assessment of each of these is more objective and appropriate.

Sacral shape

The female sacrum is considered to be shorter, wider/broad, and less anteriorly curved than the male sacrum, which, in turn, creates a larger more ovoid pelvic inlet (gynecoid as described above). The wider first sacral segment (S1) and alae breadth have consistently been the best distinguisher of females using the sacrum (Rusk & Ousley, 2016; Trotter, 1926). In females, the alae are generally as wide or wider than the promontory; and in males, the alae are narrower than the promontory (Christensen, Passalacqua, & Bartelink, 2019). The increased base breadth (S1 and large alae) creates an average sacral shape in females that is more similar to an equilateral triangle with three roughly even sides, while the male sacrum more closely resembles an isosceles triangle with the two sides being longer than the base. Males are also considered more likely to have greater than five sacral segments, which can also contribute to the overall longer appearance in some individuals (Davivongs, 1963; Rogers & Saunders, 1994). Unlike several of the other morphological traits included on this list, sacral shape and size has been extensively documented metrically (cf. Plochocki, 2011 and references within) which has further supported the visual differences mentioned here.

Visibility of the sacroiliac joint, when viewed posteriorly, also ranked in the top-10 for Rogers and Saunders (1994). The male expression includes visibility of the articular surface, while it is not visible in the female expression. This feature has received far less attention in the literature than sacral shape. Lastly, the degree of curvature is also useful for sex estimation although not included in the Rogers and Saunders' (1994) list. The greatest degree of anterior curvature in males occurs between the S2 and S4 segments (Plochocki, 2011), while females tend to have less curvature overall. The lack of a pronounced anterior curvature in females contributes to the overall gynecoid pelvis shape. Sacral curvature has also been found to be quite useful in ancestry estimation (Rusk & Ousley, 2016), and therefore, it is reasonable to assume that population differences must be considered when assessing this trait.

Subpubic concavity angle

Rogers and Saunders (1994) describe the trait as being V-shaped in males and U-shaped in females. This term is bit confusing as used by Rogers and Saunders (1994) in that Phenice (1969) specifically referred to the *concavity* as an indentation just below the symphyseal face in females (Fig. 1). The trait was either present or absent and did not exhibit a shape, per se, while the male form was simply absent. On the other hand, the subpubic *angle* is formed with the articulation of the left and right innominates and takes into account the entirety of the ischiopubic ramus (Klales et al., 2012). From the definitions provided by Rogers and Saunders (1994), it appears that they are referring to the subpubic *angle* rather than the subpubic *concavity*; however, their discussion specifically refers to this trait as Phenice's (1969) subpubic concavity.

Despite the irregularity in wording by Rogers and Saunders (1994), the two traits are interrelated. Both the subpubic concavity and subpubic angle have been extensively tested in their utility for sex estimation and shown to be both highly accurate and reliable (cf. Klales et al., 2012 and references within). A larger subpubic concavity as described by Phenice (1969) results in a larger or more obtuse subpubic angle. The subpubic angle is generally considered to be male when <90 degrees (France, 1994), yet some studies have shown that some females can be as low as 80 degrees (Decker, Davy-Jow, Ford, & Hildelink, 2011). The Klales et al. (2012) modification of Phenice (1969) changed the term to subpubic *contour* and modified the description to account for the concavity below the symphyseal face (as described by Phenice, 1969), as well as the shape of the entire ischiopubic ramus that reflects the subpubic angle. Given that both of these traits have been extensively studied and tested for validity and reliability, their use is recommended for sex estimation.

Pubis shape

The shape of the pubic bone is narrow and triangular in males, and broad and rectangular in females when viewed from the ventral or dorsal sides (Fig. 1). The proportion of the pubis as compared to the ischium varies between the sexes with females have a longer pubis length (Bruzek, 2002). This longer pubis length creates an appearance that looks "stretched" or elongated in females. The shape differences can partially be explained by the differential growth of the region during puberty (Coleman, 1969; Kerley, 1977). Differences related to ancestry and population in this feature are limited, which also make the trait useful for a broad application to sex estimation (Listi, 2010).

In females, an extra "chunk" of triangular-shaped bone is frequently found inferior-medial to the ventral arc, which in turn creates a more pronounced inferior edge (Fig. 1). This additional portion of bone creates a squared appearance in contrast to the triangular appearance commonly found in males. Overall pubis shape was incorporated into the ventral arc description in the Klales et al. (2012) revision of the Phenice (1969) method to account for these sex differences. Later, Rennie (2018) generated a new five-point ordinal scoring system that explicitly focuses on pubic bone shape; however, this method also includes the three Klales et al. (2012) revisions of Phenice's traits. Given that the new pubis shape description is combined with the Klales et al. (2012) method, which already integrates overall pubis shape, the new ordinal scale and method proposed by Rennie (2018) is somewhat redundant with the ventral arc scoring from Klales et al. (2012). However, Rennie's (2018) approach of looking at individual features and combining them for sex estimation is promising. Teasing apart the features included in the Klales et al. (2012) method—for example, separately scoring the ventral arc orientation and pubis shape rather than combined—may help us better understand the traits independently. Given the limited population differences, high accuracy rate, and low observer error of this trait, its utility as a qualitative indicator of sex is high.

Muscle markings

While muscle markings ranked high in the Rogers and Saunders's (1994) paper at seventh overall, it is unclear to which specific muscle markings they are referring. Potential muscle markings on the innominate include the gluteal lines, ischial tuberosity, pubic crest, and/or the iliac tubercle. Davivongs (1963, p. 443) broadly states that male bones are "more massive and heavier … the crests, ridges, tuberosities, and lines of attachment of muscles are more strongly marked." Given the muscle marking robusticity is heavily linked to age, body size, and activity patterns (Weiss, Corona, & Schultz, 2012 and references within), their use in sex estimation is problematic and should be avoided.

Parturition markers

These markers include dorsal pitting and the preauricular sulcus. Both are covered in extensive detail in a dedicated chapter of this volume and will not be discussed further here other than to say they should not be used as a morphological indicator of sex or parity (see Chapter 9 of this volume for more detail).

Acetabulum size and orientation

Although difficult to assess visually, the male acetabulum is, on average, larger than in females due to sexual dimorphism in body and joint size. Acetabular depth has been demonstrated to be greater in females, while width and overall dimensions are greater in males (Wang et al., 2004). In males, the acetabulum is directed laterally, whereas in females it is said be directed more anterio-laterally. The acetabular anteversion angle, which measures the anterior orientation of the acetabulum, has been found to be significantly greater in females than males (Dong, Nevelos, & Kreuzer, 2013; Wang et al., 2004). Clinical research has clearly documented that females have a larger anterior pelvic tilt angle (~3.5 degrees higher) than males, as well as a greater Q-angle/knee valgus, both of which impact the orientation of the acetabulum (Nguyen & Shultz, 2007). These differences have been attributed to differential growth during puberty and likely explain the difference in acetabular orientation between the sexes.

Simple linear measurements of maximum diameter or vertical height (cf. Kelley, 1979; Murphy, 2000) have proven easy and effective for accurate sex classification; therefore, a subjective assessment of size potentially can be eliminated. Whether acetabular dimensions such as anterior placement, depth, and diameter can be objectively scored qualitatively without the use of metrics remains to be seen. Until further research can be done to test the validity and reliability of each of these acetabular components, qualitative use of these traits should be limited or abandoned in favor of metric approaches.

Greater sciatic notch

While ranked as 12th by Rogers and Saunders (1994), the greater sciatic notch is included as one of the five traits in *Standards* (Buikstra & Ubelaker, 1994) and remains popular; therefore, a discussion of this trait is warranted. Preference for this trait likely stems from high accuracy, robustness of the element, and high survivability rates (Waldron, 1987). The greater sciatic notch, sometimes referred to as the incisura ischiadica major in older works, houses several veins and arteries and the piriformis, internal obturator, and quadratus femoris muscles (Gray, 1858). Verneau (1875) was first to notice sex differences. In males, the angel is narrower and more acute, thereby creating a deeper notch, while in females, the angle is wider and more obtuse, which in turn creates a very shallow notch (Fig. 2). According to Hanna and Washburn (1953), the notch is approximately 50 degrees in males and 75 degrees in females (as cited in Olivier, 1969). Studies examining greater sciatic notch sexual dimorphism have varied from subjective descriptive differences, such as Krogman (1962), to more objective ordinal scoring, such as Buikstra and Ubelaker (1994) and Walker (2005). Walker (2005) noted reliability issues when specimens did not exhibit morphologies at the extremes of the five-point ordinal scale and found both age-related changes and population differences in notch width. Rennie's (2018) test of the Walker (2005) ordinal scale found only moderate interobserver agreement in scoring and variable accuracy rates depending on the sample (72%–82%). Bruzek (2002) further argues that scoring the greater sciatic notch is subjective due to the influence of overall pelvis size and developmental differences in marginal structures like the ischial spine and piriform tubercle.

Perceived subjectivity in scoring this trait has resulted in many attempts to metrically capture the visible, but unreliable, sex differences. However, historically, measurements of sciatic notch depth and width have been difficult either because of damaged landmarks

Fig. 2 Left: Right innominate showing a deep "J"-shaped greater sciatic notch and a single arch chord *(dashed line)*; therefore, the composite arch is absent. Right: Right innominate showing a shallow "U"-shaped greater sciatic notch and two distinct arch chords; therefore, a composite arch is present. *(Photos courtesy of Holly Long.)*

(e.g., posterior inferior iliac spine and ischial spine) or due to repeatability issues. Bruzek (2002) attempted to modify scoring of the trait by dividing it into chords and proportions based on the work of Hanna and Washburn (1953). Analysis of the greater sciatic notch using GMM has clearly demonstrated shape differences between the sexes; however, these approaches are either not methods that can be easily applied by a user estimating sex in an unknown individual (e.g., Cline, 2015; Pretorius, Steyn, & Scholtz, 2006; Velemínska et al., 2013), or they require a large enough sample to develop reference sectioning points (e.g., González, Bernal, Perez, & Barrientos, 2007). Therefore, until more work can be done on either objectifying morphology or metrically analyzing this trait, its use for sex estimation should be limited or minimally used with caution.

Ischiopubic ramus

Like the greater sciatic notch, the ischiopubic ramus ranked low by Rogers and Saunders (1994) at 16th; however, the trait is often cited and utilized. Phenice (1969) noted that females tended to have an elevated ridge of bone just below the pubic symphysis, while the same area in males was broad and flattened (Fig. 3). Klales et al. (2012) expanded this

Fig. 3 Left: Medial aspect of the ischiopubic ramus that has an elevated ridge of bone below the symphyseal face and is narrow overall. Right: Medial aspect of the ischiopubic ramus that is broad and lacks the elevated ridge of bone below the symphyseal face.

trait to include the overall dorso-ventral width of the ascending ramus as well. In Phenice (1969), Rogers and Saunders (1994), and Klales et al. (2012), this trait was less valid and/or reliable than the ventral arc and subpubic concavity/contour traits. The Bruzek (2002) method integrates lateral aspects of the ischiopubic ramus that were first described by Novotný (1981): external eversion of the lateral border, crista phallica presence or absence, and overall robusticity. The "female expression," according to Bruzek (2002), consists of external eversion, absence of the phallic ridge, and overall gracility (Bruzek, 2002). In contrast, the "male expression," according to Bruzek (2002), consists of a lack of lateral eversion, presence of the phallic ridge, and robusticity of the ramus (Bruzek, 2002). Tests of Phenice (1969) and Klales et al. (2012) have shown the trait to be accurate and reliable (Colman et al., 2019; Kenyhercz, Fredette, Klales, & Dirkmaat, 2012; Kenyhercz, Klales, Stull, McCormick, & Cole, 2017; Walls & Klales, 2017), while tests of the Bruzek (2002) components of this skeletal region have widely varied (Blanchard, 2010; Listi & Bassett, 2006; Santos, Cuyomarc'h, Rmoutilova, & Bruzek, 2019). This trait has clear utility for sex estimation and should be included in methodological approaches that combine it with other pelvic indicators of sex (e.g., Bruzek, 2002; Klales, 2018).

Composite arch/arc compose

The utility of the composite arch for sex estimation was first proposed by Genovés (1959). The trait examines the relationship between the outline of the anterior sciatic notch chord relative to the outline of anterior segment or anterior border of the auricular surface (Fig. 2). The anterior sciatic notch chord is drawn from the ischial spine superiorly along the anterior border of the greater sciatic notch onto the blade of the ilium. In males, this chord aligns with the anterior border of the auricular surface (Fig. 2, left image), while in females, it diverges from the anterior auricular surface border (Fig. 2, right image) (WEA, 1980). Bruzek (2002, p. 161) suggests that "in females, both contours are sections of two distinct circles with different radii: the composite arch is present. Among males, both contours are a part of one circle: the composite arch is absent" (Fig. 2). To this, Bruzek (2002) added an intermediate form. One benefit of this particular trait is the high survivability of the posterior pelvis (Debono & Mafart, 2006). However, tests of the feature in different population groups have demonstrated: much higher accuracy in females (i.e., high sex bias), a high percentage of indeterminate individuals (as high as 31% in Novotoný), only moderate accuracy levels (only around ~80% in the best performing validations), and considerable variation by population studied (Bruzek, 2002; Genovés, 1959; Novotný, 1981). Due to these complicating factors, independent use of this trait for sex estimation is not recommended. Presence of the composite arch can support a sex estimation of female when present, but use of the trait to confirm male sex or without the use of other traits should be avoided.

Other features

In addition to the traits described above, limited research has indicated the utility of the following other morphological traits for sex estimation: ilium shape and orientation (Rogers & Saunders, 1994; WEA, 1980), iliac fossa depth (WEA, 1980), auricular surface elevation (Novak, Schultz, & McIntyre, 2012), aspects of the sacro-iliac joint (e.g., iliac tuberosity, postauricular space, and sulcus) (Işcan & Derrick, 1984), sacral auricular surface extension (St. Hoyme, 1984), and ischiopubic proportions (Bruzek, 2002). This last trait suggested by Bruzek is likely better evaluated metrically than visually because it is essentially a ratio. Given the exclusion of most of these traits from the Rogers and Saunders' (1994) list and many other major publications in skeletal biology, their current use is not recommended. These other traits need to be validated and tested for reliability before they can be widely applicable to sex estimation.

Employing morphological pelvic traits in sex estimation

Use of morphological traits has evolved from simple binomial (e.g., Rogers & Saunders, 1994) or presence/absence (e.g., Phenice, 1969) scoring based on the preponderance of a specific trait in one sex versus the other. This approach fails to capture the full range of human variation and excludes individuals who might be considered intermediate. To remedy this, several studies have expanded the number of categories to include an intermediate/indeterminate category (Bruzek, 2002) or have created five-point ordinal scales that are better designed to encompass the full range of human variation (e.g., Klales et al., 2012; Rennie, 2018; Walker, 2005). Very few methods, besides Klales et al. (2012) and Rennie (2018), currently exist for incorporating the results from multiple morphological traits into a statistical sex estimate, and only the former has been extensively validated. To be of utility in forensic contexts, morphological traits and methods have to include statistical estimates of error. While not necessarily true in bioarchaeological or paleoanthropological contexts, I would argue that for the sake of "good science" this same principle should apply.

At present, Bruzek (2002) is the only widely applied method that combines both morphological traits with a quasimetric approach for pelvic sex estimation. Bruzek (2002) integrated 11 components of five previously published pelvic regions (preauricular surface, greater sciatic notch, composite arch, inferior pelvis, ischiopubic proportion) to estimate sex based on a majority rule of character states (female-indeterminate-male). As described by Bruzek (2002), the ischiopubic proportions are a visual assessment of the ischiopubic ratio, so while not directly a measurement, it integrates a metric approach with the ratio. Given that many of the traits described previously may be better evaluated with metrics (e.g., obturator foramen shape or acetabulum size), a combined morphological/metric approach may be best for integrating all of the pelvic indicators of sex.

As more advanced statistical approaches are developed and more widely applied to sex estimation—for example, random forest modeling—this approach will likely gain steam and produce higher accuracy rates than when using metric or morphological traits alone (see Chapter 13 of this volume).

Limitations of morphological methods

Results have varied considerably for studies testing the reliability and validity of specific traits and methods. For example, test of the Bruzek (2002) method by Bruzek and co-authors (Bruzek, Santos, Dutailly, Murail, & Cunha, 2017; Santos et al., 2019) has produced equally as high classification results; however, other validations have varied. Some have shown comparably high overall accuracy (Listi & Bassett, 2006), while others have demonstrated considerably lower accuracy, for example, 12.3% for females and 28.2% for males in Blanchard (2010). The biggest concern with the method at present is its low reliability. Agreement (intra or inter) was only poor to fair for many of the 11 components of the five regions, and most of the individuals tested by Blanchard (2010) were categorized as indeterminate. Tests of reliability and validity of the Klales et al. (2012) method have been positive with the exception of Lesciotto and Doershuk (2018) (e.g., Klales et al., 2012; Kenyhercz et al., 2017; Klales & Cole, 2017; Stull, Kenyhercz, & L'Abbé, 2013). A review of the Lesciotto and Doershuk (2018) images (specifically their Fig. 1 on page 218) clearly indicates that they misinterpreted and misapplied the ventral arc morphology descriptions, which may explain the discrepancy in their accuracy rates compared to other studies. Most recently, Rennie (2018) integrated the Klales et al. (2012) ordinal traits, Buikstra and Ubelaker's (1994) preauricular scoring, and the Walker (2005) greater sciatic notch scale, with new ordinal scales created by Rennie for obturator foramen shape, pubic body shape, and the subpubic angle. Accuracy was very high (>90%) when using all eight traits; however, sex bias varied widely by sample (up to 17%). Reliability of traits also varied considerably, and more reliability tests need to be done. One critique of this particular method is that it moves away from a 1 to 5 scale with no assumption of sex (as in Walker, 2008 and Klales et al., 2012) back to the outdated Acsádi and Nemeskéri (1970) scale of −2 (feminine) to +2 (masculine), with zero representing indeterminate sex (see the Introduction of this volume for a more in-depth discussion on why we should be moving away from this terminology). Despite this criticism, the expansion of previously utilized traits with more objective scoring and a statistical approach to classification has the potential to greatly impact sex estimation once validated more extensively.

Population variation in the expression of morphological pelvic traits has also been explored, and the universality of methods has been tested. For example, Walker (2005) examined population differences in sciatic notch expression, and Kenyhercz et al. (2017) examined worldwide population variation in the expression of Phenice (1969) traits as described by Klales et al. (2012). Each of these studies found population differences that impacted the overall accuracy of the methods, although accuracy still remained fairly high in both (see Chapter 17 in this volume for a discussion of population variation in sexual dimorphism). Some works suggest population-specific approaches to sex estimation using the pelvis (e.g., González et al., 2007; Rennie, 2018; Walker, 2005), while others have found quite the opposite with metrics (Murail, Bruzek, Houët, & Cunha, 2005). In any case, the appropriateness of a particular method for a particular population must always be considered before being applied.

Age-related changes to traits have been studied to a lesser degree than population variation and likely need to be greatly expanded. Walker (2005) found a female bias in scoring the greater sciatic notch due to age (and population). The older an individual was, the more male-like (i.e., narrower) the sciatic notch appeared and vice versa. Likewise, age-related changes in the expression of the ventral arc have been anecdotally suggested, but not yet empirically tested (e.g., Lovell, 1989; Ubelaker & Volk, 2002). In these cases, females were incorrectly scored as males and the boney surface was much more irregular in older individuals and, therefore, more difficult to assess. Work by the author and several other authors in this volume (Stull and colleagues) is ongoing and aims to test the accuracy of these statements for the ventral arc.

Current trends

More recently, seen several noticeable trends have been seen in the use of pelvis morphology for sex estimation. Attempts have been made to metrically capture some of the shape differences between males and females via GMM (Bytheway & Ross, 2010; González, Bernal, & Perez, 2009; Klales, Vollner, & Ousley, 2009). It has been argued that GMM methods can better capture shape differences that may be difficult to quantify; however, the use of these methods can be difficult, hard to translate with casework, and requires expensive equipment to collect the data. One benefit of GMM methods is their ability to separate confounding influences of size and shape differences, which may be impossible to tease apart with traditional morphological methods. In some instances, metric assessment of specific traits may be more appropriate than trying to visually score them due to high levels of inter- and intraobserver disagreement. For example, metric and GMM assessment of differences in pelvis inlet/outlet shapes and acetabular size are much clearer. Elliptical Fourier analysis remains a good option to quantify shape differences observed with the naked eye, but that are difficult to categorize (see Caple, Byrd, & Stephan, 2017; Nawrocki et al., 2018 for a more in-depth discussion). Eventually, new computational tools will likely be developed that can make GMM approaches more user-friendly and broadly applicable. For example, Ammer, Coelho, and Cunha (2019) developed a shape simulator for the olecranon fossa of the humerus that allows the user to manipulate the outline to match their specimen. The manipulation options are based on the two principal components (PC) associated with shape (triangle vs. round) and depth (convexity vs. concavity). Once the shape is matched, in a matter of a few minutes, the program gives the user a sex probability based on linear discriminant function analysis of those PCs. This methodology could very easily be applied to the shape/size of the obturator foramen and perhaps to the depth of the acetabulum. Finally, programs such as CADOES (Coelho & Curate, 2019) and DSP2 (Murail et al., 2005) (see Chapter 15 of this volume) are also making it easier and more universal to apply metric pelvic sexing methods to unknown individuals.

Additionally, researchers have begun to validate popular traits in virtual models or create methods from virtual collections, likely due to the rapid increase in virtual anthropology and virtual autopsies. Decker et al. (2011) demonstrated that the Phenice (1969) traits can be reliably collected from digital models, while Biwasaka et al. (2009) used CT to quantify the sciatic notch shape. Colman et al. (2019) found high comparability between digital and dry bone trait scoring of the pelvis using the Klales et al. (2012) and WEA (1980) traits. Finally, and to a lesser extent, traits of the pelvis are being integrated with other skeletal morphological traits via programs like MorphoPASSE (see Chapter 16 in this volume), and sexually dimorphic adult pelvic traits are being tested for their utility in estimating sex in subadults (see Chapter 14 of this volume for more detailed information).

Recommendations

So, where does all of this leave us for morphological pelvic sex estimation? Overall, the features of the anterior pelvis are more useful for sex estimation than those of the posterior pelvis (St. Hoyme & İşcan, 1989); however, it has a lower survivability rate in paleoanthropological, bioarchaeological, and forensic contexts, which is problematic. Of those that have been extensively documented, the traits of Phenice (1969) continue to produce the highest accuracy with high reliability and should continue to be applied within the statistical framework of Klales et al. (2012) or MorphoPASSE (Klales, 2018). For the posterior pelvis, approaches like Rennie's (2018) modification of Walker (2005) is appropriate and can be broadly applied with a bit more validation. The utility of the lesser-studied traits—for example, the muscle markings and acetabulum size/orientation—needs to be further evaluated before they can be reliably used for sex estimation. And finally, several of the traits listed above—for example, the obturator foramen and pelvic inlet—might be better evaluated using simple metric approaches that can more accurately capture subtle shape differences, which in turn will improve reliability.

Conclusion

Morphological traits of the pelvis continue to be utilized for sex estimation due to their ease of use, especially when compared to metric methods, and in many cases due to their high accuracy and reliability (e.g., Phenice traits). Attempts are being made to standardize morphological trait scores (e.g., Klales et al., 2012; Rennie, 2018; Walker, 2005), integrate the use of multiple pelvic traits simultaneously (e.g., Bruzek, 2002; Klales, 2018; Rennie, 2018), validate these new approaches (e.g., Colman et al., 2019; Kenyhercz et al., 2017; Listi & Bassett, 2006), and integrate other skeletal regions with the pelvis (e.g., Klales, 2018). There has been a greater push to metricize many of the morphological pelvic traits or to at least standardize them within a statistical framework using more objective ordinal scoring. Despite the multitude of traits available for sex estimation, there have been relatively few tests of reliability and accuracy for some. We are also greatly limited in our ability to combine information from multiple traits into a probability of sex at present. In the meantime, we should be relying on the traits that have been proven valid and reliable for estimating sex from pelvis morphology.

References

Abitbol, A. (1996). The shapes of the female pelvis: Contributing factors. *Journal of Reproductive Medicine*, *41*, 242–250.

Ackerman, J. F. (1788). Dissertatio inauguralis anatomica de discrimine sexuum, Moguntiae (5 Mainz): Alef. *Ueber die körperliche Verschiedenheit*, 74–75.

Acsádi, G., & Nemeskéri, J. (1970). *History of human life span and mortality*. Budapest: Akadémiai Kiadó.

Ammer, S., Coelho, J., & Cunha, E. (2019). Outline shape analysis on the trochlear constriction and olecranon fossa of the humerus: Insights for sex estimation and a new computational tool. *Journal of Forensic Sciences*. https://doi.org/10.1111/1556-4029.14096.

Anderson, B. E. (1990). Ventral arc of the *os pubis*: Anatomical and developmental considerations. *American Journal of Physical Anthropology*, *83*, 449–458.

Baker, B. J., Dupras, T. L., & Tocheri, M. W. (2005). *Osteology of infants and children*. College Station, TX: Texas A&M University.

Bass, W. M. (1987). *Human osteology: A laboratory and field manual* (3rd ed.). Columbia, MO: Missouri Archaeological Society.

Bierry, G., Le Minor, J. M., & Schmittbuhl, M. (2010). Oval in males and triangular in females? A quantitative evaluation of sexual dimorphism in the human obturator foramen. *American Journal of Physical Anthropology*, *141*, 626–631.

Biwasaka, H., Aoki, Y., Tanijiri, T., Sato, K., Fujita, S., Yoshioka, K., et al. (2009). Analyses of sexual dimorphism of contemporary Japanese using reconstructed three-dimensional CT images—curvature of the best-fit circle of the greater sciatic notch. *Legal Medicine*, *11*, S260–S262.

Blanchard, B. K. (2010). *A study of the accuracy and reliability of sex estimation methods of the human pelvis* (unpublished master's thesis). Chico, CA: University of California.

Brown, S. (2010). Applied anatomy. In P. Bennett & C. Williamson (Eds.), *Basic science in obstetrics and gynaecology* (4th ed., pp. 57–95). United Kingdom: Elsevier.

Bruzek, J. (2002). A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology*, *117*, 157–168.

Bruzek, J., Santos, F., Dutailly, B., Murail, P., & Cunha, E. (2017). Validation and reliability of the sex estimation of the human *os coxae* using freely available DSP2 software for bioarchaeology and forensic anthropology. *American Journal of Physical Anthropology*, *164*, 440–449.

Budinoff, L. C., & Tague, R. G. (1990). Anatomical and developmental bases for the ventral arc of the human pubis. *American Journal of Physical Anthropology*, *82*, 73–79.

Buikstra, J. E., & Meilke, J. H. (1985). Demography, diet, and health. In R. I. Gilbert, Jr., & H. J. Meilke (Eds.), *The analysis of prehistoric diets* (pp. 359–422). Orlando: Academic Press.

Buikstra, J. E., & Ubelaker, D. H. (1994). *Standards for data collection from human skeletal remains: Proceedings of a seminar at the Field Museum of Natural History, organized by Jonathan Haas* (3rd ed.). Fayetteville, AR: Arkansas Archeological Survey.

Burden, B., & Simons, M. (2004). Anatomy of male and female reproduction. In C. Henderson & S. Macdonald (Eds.), *Mayes' midwifery: a textbook for midwives* (pp. 65–88). Edinburgh: Baillière Tindall.

Bytheway, J. A., & Ross, A. H. (2010). A geometric morphometric approach to sex determination of the human adult *os coxa*. *Journal of Forensic Sciences*, *55*, 859–864.

Caple, J., Byrd, J., & Stephan, C. N. (2017). Elliptical Fourier analysis: Fundamentals, applications, and value for forensic anthropology. *International Journal of Legal Medicine*, *131*(6), 1675–1690.

Christensen, A. M., Passalacqua, N. V., & Bartelink, E. J. (2019). *Current methods in forensic anthropology* (2nd ed.). United Kingdom: Elsevier.

Cleland, J. (1889). On certain distinctions of form hitherto unnoticed in human pelvis, characteristics of sex, age, and race. *Memoirs and Memoranda in Anatomy*, *1*, 95–103.

Cline, K. M. (2015). *Quantification of sex and population differences in greater sciatic notch and obturator foramen shapes* (unpublished master's thesis). Erie, PA: Mercyhurst University.

Coelho, J., & Curate, F. (2019). CADOES: An interactive machine-learning approach for sex estimation with the pelvis. *Forensic Science International*. https://doi.org/10.1016/j.forsciint.2019.109873.

Coleman, W. H. (1969). Sex differences in the growth of the human bony pelvis. *American Journal of Physical Anthropology*, *31*, 125–151.

Colman, K. L., van der Merwe, A. E., Stull, K. E., Dobbe, J. G. G., Streekstra, G. J., van Rijn, R. R., et al. (2019). The accuracy of 3D virtual bone models of the pelvis for morphological sex estimation. *International Journal of Legal Medicine*. https://doi.org/10.1007/s00414-019-02002-7.

Davivongs, V. (1963). The pelvic girdle of the Australian aborigine: Sex differences and sex determination. *American Journal of Physical Anthropology*, *21*, 443–455.

Debono, L., & Mafart, B. (2006). Sex determination from fragmented hip bones using the Bruzek method: Experience in a historic necropolis in Provence (France). *Anthropologie*, *44*(2), 167–172.

Decker, S. J., Davy-Jow, S. L., Ford, J. M., & Hildelink, D. R. (2011). Virtual determination of sex: Metric and nonmetric traits of the adult pelvis from 3D computed tomography models. *Journal of Forensic Sciences*, *56*(5), 1107–1114.

Delprete, H. (2017). Pelvic inlet shape is not as dimorphic as previously suggested. *The Anatomical Record*, *300*(4), 706–715.

Dong, N., Nevelos, J., & Kreuzer, S. (2013). Combined acetabular and femoral version angle in normal male and female populations from CT data. *Orthopaedic Proceedings*, *95-B*(Suppl. 15), 168.

Drake, R. L., Vogl, W., & Mitchell, A. W. M. (2005). *Gray's anatomy for students*. Philadelphia: Elsevier.

Fischer, B., & Mitteroecker, P. (2015). Covariation between human pelvis shape, stature, and head size alleviates the obstetric dilemma. *Proceedings of the National Academy of Sciences*, *112*(18), 5655–5660.

France, D. L. (1994). Observational and metric analysis of sex in the skeleton. In K. J. Reichs (Ed.), *Forensic osteology: Advances in the identification of human remains* (pp. 163–186). Springfield, IL: Charles C. Thomas.

Genovés, S. (1959). L'estimation des différences sexuelles dans l'oscoxal; différences métriques et différences morphologiques. *Bulletin et mémoires de la Société d'anthropologie de Paris*, *10*, 3–95 (in French).

González, P. N., Bernal, V., & Perez, S. I. (2009). Geometric morphometric approach to sex estimation of human pelvis. *Forensic Science International*, *189*, 68–74.

González, P. N., Bernal, V., Perez, S. I., & Barrientos, G. (2007). Analysis of dimorphic structures of the human pelvis: Its implications for sex estimation in samples without reference collections. *Journal of Archaeological Science*, *34*(10), 1720–1730.

Gray, H. (1858). *Gray's anatomy*. London, England: John William Parker.

Hanna, R. E., & Washburn, S. L. (1953). Determination of the sex of skeletons, as illustrated by a study of the Eskimo pelvis. *Human Biology*, *25*(1), 21–27.

Işcan, M. Y., & Derrick, K. (1984). Determination of sex from the sacroiliac joint: a visual assessment technique. *Florida Scientist*, *47*, 94–98.

Klales, A. R. (2018). MorphoPASSE: The Morphological Pelvis and Skull Sex Estimation computer program. Version 1.0. Topeka, KS: Washburn, University.

Klales, A. R., & Cole, S. J. (2017). Improving nonmetric sex classification for Hispanic individuals. *Journal Forensic Sciences*, *62*, 975–980.

Kelley, M. (1979). Sex determination with fragmented skeletal remains. *Journal of Forensic Sciences*, *24*(1), 154–158.

Kenyhercz, M. W., Fredette, S. M., Klales, A. R., & Dirkmaat, D. C. (2012). Metric and non-metric assessment of sex: Accuracy, correlation and corroboration. *Proceedings of the American Academy of Forensic Sciences*, *18*, 406–407.

Kenyhercz, M. W., Klales, A. R., Stull, K. E., McCormick, K. A., & Cole, S. J. (2017). Worldwide population variation in pelvic sexual dimorphism: A validation and recalibration of the Klales et al. method. *Forensic Science International*, *277*, 259.e1–259.e8.

Kerley, E. R. (1977). Forensic anthropology. In C. G. Tedeschi, W. G. Eckert, & L. G. Tedeschi (Eds.), *Forensic medicine* (pp. 1101–1115). Philadelphia: W.B. Saunders.

Klales, A. R., Ousley, S. D., & Vollner, J. M. (2012). A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *American Journal of Physical Anthropology*, *149*(1), 104–114.

Klales, A. R., Vollner, J. M., & Ousley, S. D. (2009). A new metric procedure for the estimation of sex and ancestry from the human innominate. *Proceedings of the American Academy Forensic Sciences*, *15*, 311.

Krogman, W. M. (1962). *The human skeleton in forensic medicine*. Springfield, IL: Charles C. Thomas.

Leong, A. (2006). Sexual dimorphism of the pelvic architecture: A struggling response to destructive and parsimonious forces by natural & mate selection. *McGill Journal of Medicine*, *9*(1), 61–66.

Lesciotto, K. M., & Doershuk, L. J. (2018). Accuracy and reliability of the Klales et al. (2012) morphoscopic pelvic sexing method. *Journal of Forensic Sciences*, *63*(1), 214–220.

Listi, G. A. (2010). The impact of racial metric variation in the pelvis in the morphological assessment of sex. *Journal of Forensic Sciences*, *55*, 1157–1161.

Listi, G. A., & Bassett, H. E. (2006). Test of an alternative method for determining sex from the *os coxae*: Applications for modern Americans. *Journal of Forensic Sciences*, *51*, 248–251.

Lovell, N. C. (1989). Test of Phenice's technique for determinant sex from the *os pubis*. *American Journal of Physical Anthropology*, *79*, 117–120.

Moerman, M. L. (1981). *A longitudinal study of growth in relation to body size and sexual di-morphism in the human pelvis* (unpublished doctoral dissertation). Ann Arbor, MI: University of Michigan.

Murail, P., Bruzek, J., Houët, F., & Cunha, E. (2005). DSP: a tool for sex diagnosis using worldwide variability in hip bone measurements. *Bulletins et mémoires de la société d'anthropologie de Paris*, *17*(3–4), 167–176.

Murphy, A. M. C. (2000). The acetabulum: Sex assessment of prehistoric New Zealand Polynesian innominates. *Forensic Science International*, *108*, 39–43.

Naňko, O., Šedý, J., & Jarolím, L. (2007). Sulcus nervi dorsalis penis/clitoridis: Its reliability as a character for sex determination of isolated human pubic bones. *Prague Medical Report*, *108*, 167–176.

Nawrocki, S. P., Latham, S. E., Gore, T., Hoffman, R. M., Byram, J. M., & Maiers, J. (2018). Using elliptical Fourier analysis to interpret complex morphological features in global populations. In K. E. Latham, E. J. Bartelink, & M. Finnegan (Eds.), *New perspectives in forensic human skeletal identification* (pp. 301–312). United Kingdom: Elsevier.

Nguyen, A., & Shultz, S. J. (2007). Sex differences in clinical measures of lower extremity alignment. *Journal of Orthopaedic & Sports Physical Therapy*, *37*(7), 389–398.

Novak, L., Schultz, J. J., & McIntyre, M. (2012). Determining sex of the posterior ilium from the Robert J. Terry and William M. Bass Collections. *Journal of Forensic Sciences*, *57*, 1155–1160.

Novotný, V. (1981). *Pohlavní rozdíly a identifikace pohlavá pínevní kosti [Sex differences and identification of sex in pelvic bone]* (unpublished doctoral dissertation). Brno, Czech Republic: Purkyně University (in Czech).

Olivier, G. (1969). Practical anthropology. *Charles C Thomas Springfield, IL,*.

Phenice, T. W. (1969). A newly developed visual method for sexing the *os pubis*. *American Journal of Physical Anthropology*, *30*, 297–302.

Plochocki, J. H. (2011). Sexual dimorphism of anterior sacral curvature. *Journal of Forensic Sciences*, *56*(1), 161–164.

Pretorius, E., Steyn, M., & Scholtz, Y. (2006). Investigation into the usability of geometric morphometric analysis in assessment of sexual dimorphism. *American Journal of Physical Anthropology*, *129*, 64–70.

Rennie, S. R. (2018). *Summary sex: A multivariate approach to sex estimation from the human pelvis* (unpublished doctoral dissertation). Liverpool, England, United Kingdom: Liverpool John Moores University.

Ridley, M. (1995). Pelvic sexual dimorphism and relative neonatal brain size really are related. *American Journal of Physical Anthropology*, *97*, 197–200.

Rogers, T. L., & Saunders, S. (1994). Accuracy of sex determination using morphological traits of the human pelvis. *Journal of Forensic Sciences*, *39*, 1047–1056.

Ruff, C. (2017). Mechanical constraints on the hominin pelvis and the "obstetrical dilemma". *The Anatomical Record*, *300*(5), 946–955.

Rusk, K. M., & Ousley, S. D. (2016). An evaluation of sex- and ancestry-specific variation in sacral size and shape using geometric morphometrics. *American Journal of Physical Anthropology*, *159*(4), 646–654.

Santos, F., Cuyomarc'h, P., Rmoutilova, R., & Bruzek, J. (2019). A method of sexing the human *os coxae* based on logistic regressions and Bruzek's nonmetric traits. *American Journal of Physical Anthropology*, *169* (3), 435–437.

St. Hoyme, L. E. (1984). Sex differentiation in the posterior pelvis. *Collegium Anthropologica*, *8*, 139–153.

St. Hoyme, L. E., & İşcan, M. Y. (1989). Determination of sex and race: accuracy and assumptions. In M. Y. İşcan & K. A. R. Kennedy (Eds.), *Reconstruction of life from the skeleton* (pp. 53–93). New York: Alan R. Liss, Inc.

Stull, K., Kenyhercz, M., & L'Abbé, E. (2013). Non-metric cranial and pelvic traits as a measure of sexual dimorphism in a modern South African population. *American Journal of Physical Anthropology*, *150*(S56), 266.

Sutherland, L. D., & Suchey, J. M. (1991). Use of the ventral arc in pubic sex determination. *Journal of Forensic Sciences*, *36*(2), 501–511.

Thoms, H. (1946). A discussion of pelvic variation and a report on the findings in 100 Negro women. *American Journal of Obstetrics & Gynecology*, *52*, 248–254.

Todd, T. W. (1921). Age changes in the pubic bone. *American Journal of Physical Anthropology*, *4*, 1–70.

Trotter, M. (1926). The sacrum and sex. *American Journal of Physical Anthropology*, *9*, 445–450.

Turner, W. M. (1885). The index of the pelvic brim as a basis of classification. *Journal of Anatomy and Physiology*, *20*, 125–143.

Ubelaker, D. H., & Volk, C. G. (2002). A test of the Phenice method for the estimation of sex. *Journal of Forensic Sciences*, *47*(1), 19–24.

Velemínska, J., Krajíček, V., Dupej, J., Gómes-Valdés, J. A., Velemínský, P., Šefčáková, A., et al. (2013). Technical note: Geometric morphometrics and sexual dimorphism of the greater sciatic notch in adults from two skeletal collections: The accuracy and reliability of sex classification. *American Journal of Physical Anthropology*, *152*, 558–565.

Verneau, R. (1875). *Le basin suivant les sexes et les races*. Paris: Parent.

Waldron, T. (1987). The relative survival of the human skeleton: implications for palaeopathology. In A. Boddington, A. N. Garland, & R. C. Janaway (Eds.), *Death, decay and reconstruction: Approaches to archaeology and forensic science* (pp. 55–64). Manchester: Manchester University Press.

Walker, P. L. (2005). Greater sciatic notch morphology: Sex, age, and population differences. *American Journal of Physical Anthropology*, *127*, 385–391.

Walker, P. L. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, *136*, 39–50.

Walls, M., & Klales, A. R. (2017). Testing the reliability of morphological sex estimation methods used in forensic anthropology. *Proceedings of the American Academy of Forensic Sciences*, *23*, 43.

Wang, S. C., Brede, C., Lange, D., Poster, C. S., Lange, A. W., Kohoyda-Inglis, C., et al. (2004). Gender differences in hip anatomy: possible implications for injury tolerance in frontal collisions. *Proceedings of the Association for the Advancement of Automotive Medicine*, *48*, 287–301.

Washburn, S. L. (1960). Tools and human evolution. *Scientific American*, *203*, 62–75.

Weiss, E., Corona, L., & Schultz, B. (2012). Sex differences in musculoskeletal stress markers: Problems with activity pattern reconstructions. *International Journal of Osteoarchaeology*, *22*, 70–80.

Workshop of European Anthropologists. (1980). Recommendations for age and sex diagnoses of skeletons. *Journal of Human Evolution*, *9*, 517–549.

CHAPTER 7

Adult sex estimation from cranial morphological traits

Heather M. Garvin
Department of Anatomy, Des Moines University, Des Moines, IA, United States

If you look at a person's face, even in the absence of makeup or facial hair, you can generally discern whether that individual is biologically male or female. Much of this is due to underlying bony cranial morphology, and it is these sex traits that allow forensic and biological anthropologists to successfully differentiate male from female crania. Sexually dimorphic cranial traits, such as the browridge, are sometimes referred to as morphological traits, nonmetric traits, or discrete cranial traits. These terms are used to distinguish these cranial features from overall cranial morphology (i.e., shape), which is commonly analyzed metrically (e.g., in the computer program FORDISC). Instead, these morphological traits refer to specific features on the cranium and mandible, localized regions where sex differences are observed and can be assessed independently of other features. Note also that, although some of the traits fall on the mandible and thus skull traits would be the most appropriate term, they are typically referred to as "cranial traits." How this misnomer developed is unclear, but it may be for simplification or as an obvious counterpart to postcranial traits. The author debated using "skull traits" here instead, but fell back on the conventional terminology: (1) admittingly, it feels more natural and (2) so that there is a clear connection between what is discussed here and in other literature, as the term "cranial traits" can be found in most keywords and search terms.

Historically, qualitative analyses were used to assess sex from these traits, as their dimorphic shape is difficult to adequately capture metrically and they tend to lack true bony (Type I) landmarks. The terms nonmetric traits or discrete cranial traits are somewhat outdated and do not accurately reflect the traits of interest (see Introduction of this volume). As we will see, most of the sexually dimorphic cranial traits are not truly discrete or categorical traits, but have a gradient of expressions across the male and female spectrum (although practitioners are sometimes forced to place them in discrete categories in some methods). And as technology continues to advance, more quantitative methods are being derived to analyze these traditionally "nonmetric" traits.

Traditional morphological trait analyses

It is generally common knowledge that males tend to be more robust than females, and this pattern continues in the skull. When presented with a female and a male skull

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00007-9

side-by-side, even an untrained eye can generally assign the correct sex to each most of the time. Compared to females, males tend to have more robust muscle attachment sites (e.g., nuchal crests, mastoid processes, temporal lines, zygomatic extensions, gonial eversion), more projecting browridges and glabellar regions, more vertical foreheads, reduced parietal and frontal bossing, and more squared mandibles with prominent mental eminences (Figs. 1 and 2). Sex estimation becomes more difficult without the pairwise comparison and with the presence of ambiguous trait expressions and the effects of population variation. More formal analyses of sexually dimorphic cranial traits are documented as early as Broca (1875) and Acsádi and Nemeskéri (1970) were among the first to create trait scoring methods for sex estimation.

The full list of cranial and mandibular sexually dimorphic traits is relatively long. Krogman (1955) presented a list of 13 cranial sex traits. Rogers (2005) presents Krogman's list in her Table 1, along with information regarding which of the six popular forensic anthropology textbooks recommend each of the traits. Rogers then added an additional four traits and conducted her own analysis, scoring the traits binomially on a Canadian sample, and found that of the 17 traits, nasal aperture size/shape, zygomatic extension, malar

Fig. 1 Photograph of a cranium displaying more robust, male-like cranial features (left) compared to one displaying more gracile, female-like traits (right).

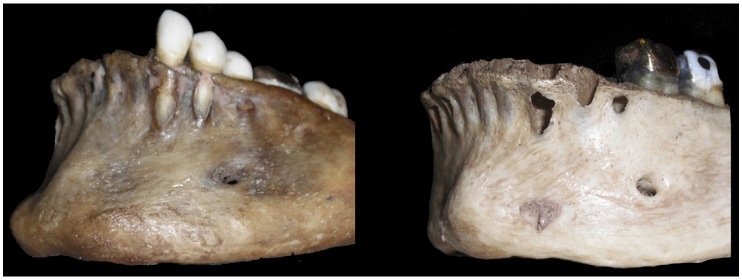


Fig. 2 Photograph of a more characteristically male mental eminence (left) with projecting lateral tubercles, compared to a more characteristically female mental eminence (right) displaying a pointed mental eminence lacking lateral extension.

Table 1 Summary of cranial morphological sex estimation techniques discussed in the chapter.

Summary of skull morphology sex estimation techniques	
Scoring techniques	
Walker (2008)	Five-scale ordinal scoring of five most popular traits Also published in Buikstra and Ubelaker (1994)
Rogers (2005)	Binomial scoring of 17 traits Includes Krogman's (1955) 13 traits; also used/modified by Williams and Rogers (2006)
Walrath, Turner, and Bruzek (2004)	Five-scale ordinal scoring of 10 traits, with character weights Is a combination of traits from Ferembach (1980) and Walker (2008) (from Buikstra & Ubelaker, 1994)
Sex estimation techniques	
Walker (2008)	Logistic discriminant functions (combined US white, US black and English for modern sample, separate Native American sample); also provides univariate frequencies and probabilities
Garvin, Sholts, and Mosca (2014)	Pooled, population, and sample-specific discriminant functions for US white, US black, Nubian, and Native American; also provides univariate frequencies and probabilities
Stevenson, Mahoney, Walker, and Everson (2009)	CHAID decision trees using Walker (2008) data; decision trees for pooled US white, US black, and English sample, as well as a separate US white tree
Langley, Dudzik, and Cloutier (2018)	Decision tree using glabella, mastoid, and zygomatic extension from US whites
Ferembach (1980)	Multiplies specified trait weight to ordinal score (scale of -2 to $+2$), such that a positive score indicates male, and negative score indicates female; based on their vague trait descriptions so that can be interpreted as being population-specific; explained in Walrath et al. (2004)

size/rugosity, and the supraorbital ridge were the most useful for sex estimation. When Williams and Rogers (2006) similarly analyzed a set of US skulls from the William M. Bass Donated collection, their trait rankings shifted place (e.g., mastoid process was ranked among the most reliable), perhaps reflecting population variation in trait expressions between the modern US sample and historical Canadian sample. Through both of their studies, however, the supraorbital ridge remained the highest ranked in terms of accuracy. Palate size/shape, orbit size/shape, frontal and parietal eminences, occipital condyle size, and tooth size were listed among the worst of the traits (Rogers, 2005; Williams & Rogers, 2006).

Some anthropologists may favor certain traits presented by Krogman (1955) and Rogers (2005) over others based on their experiences (see Chapter 2 of this volume).

The binomial nature of Roger's method basically leaves the practitioner to decide whether each trait appears more male/robust or more female/gracile in morphology. Others may incorporate a number of these traits, either consciously or subconsciously, when analyzing the overall "gestalt" of the skull for sex estimation. Five particular morphology traits, however, have received the most attention and are perpetuated in recent literature: supraorbital ridge/glabella, supraorbital margin, mastoid process, nuchal crest, and mental eminence. The popularity of these five traits is likely due, in part, to their supposedly high reported accuracy rates (see below for discussion), but is also likely related to their publication in Buikstra and Ubelaker's (1994) *Standards for Data Collection from Human Skeletal Remains*. *Standards* quickly became a staple text for practicing forensic anthropologists and bioarchaeologists as well as their students. Contained within this popular text are instructions for scoring these five morphological traits, complete with line drawings and written descriptions.

The method presented in *Standards* entails scoring each of the five skull features on an ordinal scale from 1 to 5 based on a series of line drawings and written descriptions. The lower the score, the more female/gracile the morphology; the higher the score, the more male/robust the morphology. The scoring method was based off of Acsádi and Nemeskéri's (1970). Buikstra and Ubelaker (1994) did provide vague instructions for estimating sex, suggesting that practitioners score all cranial as well as pelvic traits and determine the overall individual score: 1 = female, 2 = probable female, 3 = ambiguous sex, 4 = probable male, or 5 = male. It is unclear whether the final individual score would be an average of the trait scores, modal trait score, or subjectively determined based on overall impressions. There is no suggestion that certain traits may be better or should hold more weight than others.

In 2008, Walker republished the cranial trait scoring method (he actually states that he was the one who created and contributed the scoring method published in *Standards*) along with frequency statistics and discriminant functions for sex estimation. Walker's sample consisted of scores from 304 US white, US black, and English skulls and from 135 Native American skulls. He provides score frequencies for males and females and sex probabilities associated with each univariate score. He also presents logistic discriminant functions, allowing practitioners to input their trait scores into one of the provided equations to more objectively arrive at a final sex estimate. This increases the statistical rigor, although subjectivity in the scoring method remains. The exclusion of some variables from the equations and variable coefficient weights suggests that certain traits are more useful than others. Univariate analyses indicate that glabella performs the best (82.6% correct classification), followed by the mastoid process (78.6%). Walker's (2008) highest performing discriminant function included glabella, mastoid, and mental eminence scores with a resultant 88.4% correct classification for males and 86.4% for females.

Walker (2008) method validation

There are two aspects of the "Walker method" that require validation—the first being the trait scoring method, and the second being the sex estimation equations. Despite the popularity of the Walker traits, there are relatively few validation studies looking at either of these steps.

Validation of Walker scoring method

In 2004, Walrath et al. assessed the reliability in scoring of a number of cranial traits, including the five traits that are part of the Walker (2008) method. Note that they were using criteria from both Ferembach (1980) and Buikstra and Ubelaker (1994) when scoring traits, so it is not an ideal test of the specific Walker method, but still provides information regarding trait reliability. As they note, the Ferembach (1980) article provides brief trait descriptions and uses a scale of −2 to +2 mirroring Acsádi and Nemeskéri (1970) (instead of 1–5), but provides weights for the different traits and a method of sex calculation $[(\sum(\text{score} \times \text{weight}))/\sum\text{weight}]$. The scoring materials in Buikstra and Ubelaker (1994) are more detailed, potentially decreasing subjectivity. Thus, they used both sets of materials and the Ferembach (1980) estimation method. They found that superciliary arches (i.e., browridges), glabella, external occipital protuberance (i.e., nuchal), and mastoid process had high levels of interobserver reliability.

Walker (2008) did provide reports of inter- and intraobserver agreement based on a subsample of 10 skulls analyzed by 20 observers of various experience levels. All observers, however, were able to ask Walker questions regarding the scoring method, thus eliminating some of the blindness of the method. Walker reported that all observers fell within one score of the modal value 96% of the time. The intraobserver analysis was conducted by himself, in which he found that 99.5% of the scores from five trials fell within one score of the modal value. From his reports, it remains unclear whether the difference of one score would affect sex estimation results and what bias his involvement in the intra- and interobserver analyses created. Furthermore, he mentioned specifically selecting the 10 test skulls to represent the overall trait variation, which may have avoided skulls with some of the more ambiguous trait expressions.

At present, the only external study directly testing the reliability of Walker's cranial trait scoring method (2008) and discriminant functions is that of Lewis and Garvin (2016). Lewis and Garvin each scored the five cranial traits using the Walker (2008) materials on a sample of 135 19th- and 20th-century US white and US black skulls (comparable to those included in Walker's study). They both also rescored the US black sample (with 1 week between trials for Lewis and 2 years between trials for Garvin). In general, their intraobserver results mirror Walker's in terms of consistently being within one score. Intraobserver analyses produced moderate to strong agreement in all five traits for both observers, with the exception of Lewis' scores for the mental eminence. Interobserver analyses indicated strong agreement between all traits, except for the mental eminence. There were, however, some statistical systematic bias between observers, with one observer systematically scoring the mastoid and glabella traits higher in the white sample and the other scoring the nuchal crest and mental eminence higher in the black sample.

Lewis and Garvin (2016) also had three-dimensional surface scans of the black sample of skulls. Using the method published by Garvin and Ruff (2012), they used semilandmarks across the brow and chin regions to quantify the brow and chin shapes of the sample. They then compared the principal component and resultant discriminant function scores from those morphometric analyses to the ordinal trait scores to evaluate how well their assigned trait scores were capturing actual shape variation. For the browridge, significant moderate correlations were obtained between the brow discriminant function scores and visually assigned ordinal scores (*r* ranging from 0.63 to 0.72, depending on observer and trial). A discriminant function performed solely on the morphometric data returned a cross-validated correct classification rate of 83.3%. There was less concordance between the morphometric and ordinal score data for the mental eminence. There was not a significant correlation between the values for Lewis, and the *r*-values for Garvin's scores ranged from 0.27 to 0.34. The discriminant function performed solely on the mental eminence morphometric data was only able to correctly sex 71.7% of the sample (cross-validated).

Between the intra- and interobserver analyses, and the correlation analyses with morphometric data, a pattern emerges in which the mental eminence does not appear to be a strong variable for sex estimation. The authors note that Walker's scoring method for the mental eminence combines multiple traits (e.g., area of involvement and degree of projection) into a single gradient, when in their experience the variables did not necessarily co-vary, leaving the observer to subjectively choose how to compensate for the mixed signals. See Fig. 3 for an example of variation in mental eminence morphology. It was also pointed out that one observer scored most of the mental eminences between 2 and 4, effectively reducing the gradient to a three-point scale. While it appears apparent that there are issues to mental eminence scoring procedure, even the morphometric analyses suggest that the mental eminence can at best only correctly classify about 72% of individuals. In a separate study, Langley et al. (2018) similarly found that the mental eminence did not contribute significantly to their decision tree sex estimation procedures and had low interobserver agreement; consequently, they suggest avoiding the use of this feature.

Given these challenges, it is surprising that the mental eminence is included in most of the discriminant functions presented by Walker (2008). Although Garvin et al. (2014) also noted issues with mental eminence, when they performed a stepwise discriminant function, the mental eminence was retained in the equations for US blacks (but not US whites), their archaeological samples (Native American and Nubian), and their

Fig. 3 Photographic examples of varying chin (mental eminence) morphologies.

overall pooled sample equations. The retention in the discriminant function analyses suggests that it was contributing something to the discrimination between sexes in those groups; however, interobserver issues in scoring this feature may obscure any discriminating potential. Perhaps, newly derived scoring methods focusing on individual aspects of the mental eminence (e.g., scoring area involved and lateral tubercle projection separately) may provide more reliable results.

There is a consensus among studies that the best variable is the glabella/supraorbital ridge followed by the mastoid process (Abdel Fatah, Shirley, Jantz, & Mahfouz, 2014; Garvin et al., 2014; Garvin & Klales, 2018; Garvin & Ruff, 2012; Stevenson et al., 2009; Walker, 2008; Williams & Rogers, 2006). Both consistently displayed high intra- and interobserver agreement levels, as well as high accuracy levels. Garvin et al. (2014) report an overall correct classification rate of 79% (cross-validated) if only using glabella, and 76% (cross-validated) if only using the mastoid process on a pooled large ($n = 499$) and diverse sample (US whites, US blacks, medieval Nubian, and plains Native American). Accuracy rates generally increased if analyzed within each of the population groups. The higher accuracy rates for glabella and mastoid process are likely a consequence of both greater dimorphism in these traits as well as better scoring standards (line drawings and descriptions) than what is present for other traits.

Interobserver analyses conducted by Lewis and Garvin (2016) and Walls, Klales, Lesciotto, Gocha, and Garvin (2018) also suggest that although the interobserver agreement was generally acceptable (with the exception of the mental eminence), experience does play a role in cranial trait scoring. Observers with more experience generally had higher intraobserver agreement and higher correct sex classification rates utilizing their scores. This emphasizes the subjective nature of assigning ordinal scores to traits. Furthermore, although statistical analyses generally indicate acceptable intra- and interobserver levels of agreement in trait scores, studies have shown that even a difference of a single score (i.e., scoring a trait a 2 versus a 3) can change the sex outcome (Lewis & Garvin, 2016). Lewis and Garvin (2016) report a change in sex outcomes up to 36.6% in intraobserver trials and up to 44.9% between observers when Walker's (2008) discriminant functions were applied. Thus, even slight differences in scores impact sex classification accuracy.

Validation of Walker discriminant functions

The best performing "modern" discriminant function presented by Walker includes scores from glabella, mastoid process, and mental eminence, and he reports a correct classification rate of 88.4% for males and 86.4% for females. The equation was derived from a pooled sample of US whites, US blacks, and English skulls. When Lewis and Garvin (2016) applied this equation to their US black sample, depending on the observer, trial accuracy ranged from 56.7% to 73.3% for males and 76.7% to 90.0% for females. These classification rates are generally lower than those reported by Walker, and illustrate a strong sex bias with higher classifications in females. Accuracy rates were higher when applied to their US white sample, with the male percent correct ranging from 73.0% to 83.3% and the female percent correct at 89.7%. A sex bias toward females was still noted. Accuracy rates for the other Walker equations were highly variable and dependent on the observer and trial. Walls et al. (2018) report an overall correct classification of 73.5% for expert observers, 61.4% for experienced observers, and 70.7% for inexperienced observers when using Walker's first discriminant function.

Garvin and Klales (2018) specifically tested Walker's second equation that utilizes only glabella and mastoid (and not the problematic mental eminence). For this equation, they obtained only a 59.5% accuracy rate for US white females, 80% for US black females, 97.5% for US white males, and 86.9% for US black males. This equates to an overall percent correct of 68.3% for females, 93.0% for males, and 80.8% overall. Again, sex bias is evident, although it reversed with more specimens being classified as males.

These studies suggest that sex classification rates are generally lower than those reported by Walker (2008). Differences in classifications go back to the subjectivity of the scoring method. In order for the equations to perform optimally, the traits must be scored in the same manner as they were in the sample from which the equations were derived. In other words, practitioners would need to score specimens exactly as Walker did originally. Given the documented subjectivity in scoring traits (particularly with experience level), this is not likely to happen. Systematic differences between observers, in particular, could lead to large sex biases.

There is actually an argument to be made that, because of the inherent subjectivity in the trait scoring method, it may be better to utilize trait scores in more flexible/subjective sexing methods instead of in equations that provide an illusion of concrete objectivity. Walrath et al. (2004) actually state that the subjectivity of Ferembach's (1980) scoring system is an advantage, as a practitioner can mentally revise the gradient scale to correspond with the variation observed within the specific population they are working with. What constitutes a marked external occipital protuberance in one population group may differ from the next. Although there is a general push in forensic science toward more standardization and objectivity given the guidelines presented in Daubert (1993) and the National Academy of Sciences (2009) report, many practitioners continue to perform merely an overall visual assessment of the skull for sex estimation, many times referred to as the "gestalt" method. Such a method is subjective, dependent on the observer and their exposure to population at hand, and yet can be highly accurate when performed by experienced observers. Lewis and Garvin (2016) tested such a "gestalt" method, and experienced observers obtained classification rates higher than any of the results from the Walker equations (93.9% for US white males and 92.6% for US white females). The results from less experienced observers, however, were not as impressive (64.7% for US white males and 85.7% for US white females).

Effects of other variables on cranial traits and sex estimation

Population variation

Populations vary in their degree of expression of morphological cranial traits as well as their degree of sexual dimorphism in the traits. In other words, some populations may display more male/robust or more female/gracile traits, and some will display greater separation between the sexes in certain traits. Furthermore, these population effects can be trait-specific. One population may display greater dimorphism in glabella, while another in the mastoid process. Walker (2008) discusses population variation in trait expression, noting, in particular, that the English sample (which was included in the pooled modern sample for discriminant functions) was overall more gracile than the US white and US black samples. Among other group differences, he notes that the US black males had significantly larger mental eminences than US white males. In contrast, US white males had more prominent glabellar projections and more rounded orbital margins than the US black males. Garvin and Ruff's (2012) morphometric analysis of the browridge and chin support these US patterns. Despite statistical differences between groups in various traits, Walker (2008) reported that coding for the various modern population groupings did not significantly improve the classification rates; and as such, he grouped them for discriminant functions.

Garvin et al. (2014) found significant effects of population on all five of the Walker traits and significant interactions between sex and population in nuchal, glabella, and orbital margin traits. They also found that US whites displayed the most robust and dimorphic glabellar regions, consistent with Walker (2008). Although their own discriminant function (which retained all variables except the nuchal scores) performed well on the pooled sample (85% cross-validated), population and sample-specific equations ranged in accuracy from 74% to 94% and retained different variables and variable coefficients when maximizing discrimination between the sexes in each of those groups. Although pooled equations may provide overall decent classification rates, there is likely to be population bias, with males in more gracile populations frequently being misclassified as females, and females in more robust populations as males. Consequently, if the population can be estimated reliably, population-specific methods/equations would provide most accurate results. For those populations with less sexual dimorphism, accuracy levels would be expected to be lower than those that have greater sexual dimorphism, but at least the methods would reflect true population variation. In forensics, it's not about getting the best overall numerical value; it is about best representing the true group variation so that the chances of correctly identifying a single specimen increases.

Age

In a 1995 article, Walker suggested that supraorbital robusticity increases in males over the age of 30, and in females over the age of 45 years. In his 2008 study, he noted a significant association between birth year and age with trait scores in his model analyses. He found, however, that adding age parameters did not appreciably improve the models. Garvin et al. (2014) tested for the effects of age using Jonchkeere-Terpstra trend tests and Spearman rank correlation analyses and found an overall positive relationship between cranial trait scores and age (overall and within each sex), although the percentage of trait variance explained by age was <13%. When populations and samples were analyzed independently, the significant correlations between age and traits varied greatly, with no obvious trends. Overall, age does not appear to have a large enough impact on cranial trait scores to necessitate their inclusion in sex estimation methods, especially given that the precise age of the decedent is unknown in most forensic cases. Morphometric methods that do not involve pigeonholing trait variations into a scale of five scores may be more sensitive to age-related changes, and such research should consider including age as a potential variable.

It is important to note that the use of morphological traits in subadults for sex estimation is not recommended. Sexual differentiation in trait expressions is believed to begin with hormonal changes around puberty; and thus, all of the methods described in this chapter are solely for adult sex estimation. See Chapter 14 for more information about subadult sex estimation.

Body size

Garvin et al. (2014) also looked into the relationship between body size and cranial trait scores. They used femoral length and femoral head diameters as a proxy for stature and body mass, respectively. They found an overall increase in nuchal, glabella, and mastoid scores with both femoral variables in both sexes using the Jonchkheere-Terpstra tests, and significant correlations between body size variables and most of the cranial traits, although r-squared values were <0.14. Again, however, if within-group analyses were conducted, no specific patterns in significant relationships were observed. Horbaly, Kenyhercz, Hubbe, and Steadman (2018) found similar results when they analyzed the relationship between cranial traits and documented statures and body masses. Overall, they found significant correlations between trait scores and body size parameters (and a stronger relationship with stature than body mass), but those relationships that were statistically significant were weak. As with age, the impact of body size on cranial trait scores is not great enough to warrant inclusion in forensic estimation methods.

Function

Although the five Walker traits are generally grouped together, we must remind ourselves that each one reflects a unique skeletal adaptation and has a different function. The mastoid process is fairly straightforward; it is an attachment site for a number of muscles, including the sternocleidomastoid, splenius capitis, the posterior belly of the digastric and longissimus capitis muscles. Thus, it can be inferred that where larger muscles are involved, a larger attachment site is needed. Sex differences in mastoid process size are likely the result of differences in body size, muscle mass, and activity levels in males and females. This is also true of the nuchal crest/external occipital protuberance. The superior nuchal line provides an attachment site for the trapezius, occipitalis, and sand splenius capitis muscles, while the external occipital protuberance is the attachment site for the nuchal ligament.

The function of the browridge (supraorbital ridges, glabella) is less understood. Some have suggested that it has biomechanical functions related to mastication (Endo, 1970; Oyen, Rice, & Samuel Cannon, 1979; Russell et al., 1985), although there is little empirical support (Baab, Freidline, Wang, & Hanson, 2010; Bernal, Perez, & Gonzalez, 2006; Hylander, Johnson, & Picq, 1991; Hylander, Picq, & Johnson, 1991). Others support a structural hypothesis, suggesting that the browridge forms a structural bridge between the neurocranium and viscerocranium as the two regions develop independently (Fiscella & Smith, 2006; Moss & Young, 1960; Ravosa, 1988). Still others suggest that it is related to the overall craniofacial size (Ravosa, 1988). There are even some who suggest that browridges evolved to keep the sun, sweat, or rain out of our eyes on the African plains (Boule & Vallois, 1957; Davies, 1972; Guthrie, 1970; Krantz, 1973; Kurten, 1979), or snake venom out of our eyes (Davies, 1972). The most recent hypothesis suggests that browridges evolved because we like to punch each other in the face (Carrier & Morgan, 2015) or for subtle affiliative emotions (Godinho, Spikins, & O'Higgins, 2018). Basically, we have yet to figure out why we have browridges or why their size and shape vary among populations. The sex differences observed, however, are likely related to hormone (e.g., testosterone) differences between the sexes, and perhaps sexual selection is involved as well. Little research has focused on the supraorbital margin morphology, although Garvin et al. (2014) did discover a significant correlation between orbital margin scores and glabella scores, and hypothesized that rounding of the supraorbital margin may be related to the overall inflation of the area with more prominent supraorbital ridges.

The mental eminence, or chin, is a similarly enigmatic feature (Daegling, 1993). Again, a biomechanical function related to masticatory stress has been proposed but lacks empirical support. Others have suggested that the chin, which is fairly unique to *Homo sapiens*, formed as a byproduct of a receding dental arcade (with our smaller teeth) (Riesenfeld, 1969; Weidenreich, 1936). Sexual selection has also been proposed for the development and variation in the mental eminence (Hershkovitz, 1970; Lieberman, 1995; Penton-Voak & Chen, 2004). Garvin and Ruff (2012) actually found that there are more population differences in the shape of the mental eminence than sex differences.

Given that we don't fully understand these skeletal adaptations, why they evolved, or what function(s) they hold, it is difficult to explain variations in trait morphologies. Obviously, sex hormones likely play a role in all of these traits, but it is not the only factor (or else all of these traits would co-vary tightly with one another, when in fact we see a lot of within-individual variation in trait expressions). Investigating the factors contributing to trait variations is difficult when the traits themselves are hard to quantify. Geometric morphometric analyses provide a unique tool to begin evaluating factors contributing to trait and population variation. For example, Garvin and Ruff (2012) found that browridge *shape* actually varies among different populations, not just prominence. US whites tend to display more "drooping" brows, at times producing a "W"-shaped prominence at glabella, while US blacks displayed more "V"-shaped projections, angled higher above the orbit margins. Such shape variations can get lost when the focus is on a simple gradient five-point scale, and may be important to figure out the variables contributing to sex and population differences.

Applications of different methods

A few studies have applied Walker's scoring technique to methods other than discriminant function analysis to estimate sex. Stevenson et al. (2009) actually used the same cranial score data that were collected and used by Walker, but applied a Chi-square automatic interaction detection (CHAID) procedure to produce a decision tree for sex estimation. A decision tree is basically a dichotomy key in which the user starts with a specified trait and, based on that trait score, then follows one of the two pathways to the next node where, based on the next trait score, they yet again follow one of the two paths to the next node. The nodes and the levels at which a "stop" is encountered (where final sex is estimated) are determined through computational iterations to optimize sex discrimination. All of the trees produced by Stevenson et al. (2009) included some combinations of glabella, mastoid, and mental eminence scores, consistent with Walker's (2008) discriminant function (and not surprising since he is using Walkers data). The overall accuracy of the acceptable trees for the pooled modern sample hovered around 85%. They did also produce population-specific trees, but found that none of the trees produced results with >75% accuracy and <5% sex bias for the US blacks. The inability to produce an acceptable US black tree supports claims of population variation in traits and dimorphism and is consistent with the lower accuracy rates obtained by Lewis and Garvin (2016) for the US black sample. When Garvin and Klales (2018) applied Stevenson et al.'s European-American tree (which included only the glabella and mastoid), they achieved a 96.2% accuracy for US white females and an 80.2% accuracy for US white males.

Langley et al. (2018) also took a decision tree approach. They collected scores for the five typical cranial traits as well as zygomatic extension (suprameatal crest). They produced a decision tree that utilized glabella, mastoid, and zygomatic scores and reported correct sex classifications between 94% and 96%. When Garvin and Klales (2018) tested the Langley et al. decision tree, they found it only correctly classified 71.5% of their pooled samples, with a very strong sex bias (94.2% correct for females and 49.3% for males). They also reported some subjectivity and interobserver issues in scoring the zygomatic extension, which may have played a role in the lower classification rates.

Garvin and Klales (2018) bring up the fact that extreme trait scores are often more reliable than intermediate scores. In general, there is more agreement between observers on trait expressions on the extremes of the scales, and these extreme scores are usually more diagnostic. A cranium with a complete lack of browridges is more likely to be female than one with slight or moderate browridges. They found that individuals with a glabella score of 1 had a 94% chance of being female, while those scored a 4 or a 5 had a 98% probability of being male. Those with a mastoid score of 5 or a supraorbital margin score of 5 had a 100% probability of being male. A nuchal score of 5 indicates a 94% probability of being male. If a skull exhibited more than one of these extreme trait scores, sex was pretty much guaranteed. A similar pattern was observed when one looks at

Stevenson et al.'s (2009) European-American decision tree (their Fig. 2). At their first node, a glabella score of 1 or 2 indicated a 91% probability of being female, and analyses stopped at that level. A glabella score of 4 or 5 indicated a 95% probability of being male, with analyses stopped there. Only those with a glabella score of 3 were actually directed to the next node in order to score the mastoid process before estimating sex. These univariate analyses suggest that including more variables does not necessarily increase accuracy. The presence of an extreme trait score may be enough to accurately predict sex, before muddling the analyses with additional and potentially more ambiguous traits. Walker (2008) and Garvin et al. (2014) also provide trait score frequencies and probabilities in their tables, which support the notion that extreme trait scores provide more confident sex assessments and can be used for univariate analysis.

Quantifying the traits

Given the aforementioned issues with subjectivity in cranial trait scoring, some researchers have begun to develop morphometric techniques to better capture and quantify the size and shape variations of individual cranial features. Gonzalez, Bernal, and Perez (2011) used semilandmarks along a curve for glabella and mastoid (as well as frontal and zygomatic processes) to capture sexual shape variation. Their best discriminant analysis returned a 78% correct sex classification. Garvin and Ruff (2012) used semilandmarks along the surface of browridge and chin to capture the full trait morphology. Their discriminant function analyses returned a 79.8% correct sex classification (leave-one-out cross-validation) for the browridge and only 62.2% for the mental eminence. These methods would be too time-consuming to apply for an individual forensic case, but provide background research into the levels of dimorphism that are expected for these traits and means for investigating relationships into other variables. If a computer with all available shape information can only obtain an 80% correct sex classification, can we expect other methods, such as visual assessment, to perform above that? And if they do, could it be because during visual assessments we are incorporating details into our evaluation other than that single trait (e.g., bias of overall cranial size, etc.)?

Recently, Nikita and Michopoulou (2018) went a step further and produced a semi-automated method to evaluate frontal outlines (including glabella), mastoid outlines, and nuchal outlines from lateral photographs of crania. The outline data were then run through an Excel macro that calculated a set of discriminant variables (such as centroid size, mastoid process length and width, differences in glabella coordinates associated with projection, etc.), which were then automatically run through the logistic regression models. They reported correct sex classification between 80% and 90% depending on the sample (they used a Greek and Cretan sample). This provides a method to objectively capture these shape variables without adding too much time to the analysis, and it will be interesting to see how subsequent validation studies perform.

Murphy and Garvin (2018) used complete cranial outlines to assess sex and ancestry differences. The sex differences captured by the cranial outlines (contours) mirrored qualitative descriptions of sex differences, picking up differences in mastoid shape, projection, and volume, nuchal morphology, glabellar projection, a frontal and parietal bossing. Overall sex classification, however, was relatively low (70.2% cross-validated), although ancestry classification (92.4% cross-validated) was higher, and 66.7% (cross-validated) were correctly assigned their specific ancestry + sex group, which is significant given that the a priori classification rate is only 25%. Interestingly, posterior vault outlines almost performed as well as lateral cranial outlines, with an accuracy rate of 70.1% (cross-validated). Caple, Byrd, and Stephan (2018) performed similar outline analyses, but incorporated their data and method into a freely available script that can be run using the statistical program R. This allows practitioners to quickly analyze their cranial outlines as compared to a large reference sample. Their results mirrored those of Murphy and Garvin (2018), suggesting that outline analyses that incorporate the shape of multiple cranial traits as well as overall cranial morphology may provide a promising avenue for future methods.

Abdel Fatah et al. (2014) went one step further, creating an automated method that compares a full 3D model of a cranium (from CT scans) to an atlas model to estimate sex. Although it uses the entire cranium, it is picking up morphological differences in the glabellar region, inclination of the frontal bone, cranial base flexion, as well as size differences in bizygomatic breadth, maximum cranial length, cranial base length, and mastoid height. They report more than a 95% accuracy (cross-validated). Although the method provides rapid results (if a 3D CT model of the cranium in question is already available), it has not yet been made into a freely available program for others to implement. At present, CT scans also remain relatively expensive and are not commonly accessible by most forensic anthropologists. Methods such as Abdel Fatah et al.'s (2014) automated atlas method are likely to become more common as technology continues to advance and CT scans and 3D models become more commonplace.

Conclusion

Based on published research, it is evident that cranial morphological scoring methods have subjectivity issues. When forcing the gradient (continuous) variation observed in cranial traits into discrete ordinal categories, there is going to be some inter- and even intraobserver error. Although these values are generally statistically acceptable, studies have shown that they can dramatically impact overall sex estimation. Experience plays a major role, with more experienced practitioners displaying better intraobserver agreement and higher sex classification rates, likely because of their increased exposure to the amount of skeletal variation that can be expected of the population at hand. Of the cranial traits, glabella/supraorbital ridges and the mastoid process are the most reliable in terms of better intra- and interobserver agreement and higher classification rates, while the mental eminence has proven to be problematic.

With such issues, why do practitioners continue using cranial morphological traits in sex estimation? Ultimately, because the methods still work. Accuracy rates across the board range from about 80% to 90%. It may not be as accurate as pelvic methods, but can complement those analyses or are necessary in the absence of pelvic features. Morphological trait methods are also simple and can be completed in a matter of a few minutes, expediting analyses. They do not require any specialized equipment or advanced technology and can be completed on fragmentary or otherwise modified remains.

In terms of which specific method to utilize, the verdict is still out. Various studies report different accuracy rates, although almost all suggest that glabella and mastoid process are the best features. The author believes that it is hard to gauge which analytical method is the most appropriate when the scoring method retains these inherent issues. How can you determine if one statistical method or equation is better than another when the ordinal scores used to develop them have subjectivity concerns? There needs to be a better consensus on scoring methods before the most appropriate application can be decided. This may require new scoring methods or materials (images and descriptions) and, in some cases, focusing on individual trait characteristics instead of grouping multiple characteristics into a single gradient. Beyond trait scoring issues, practitioners are likely to implement the methods they were trained on or those that they are most familiar with and have previously been successful in their casework. Method decisions should also be made based on the most similar reference populations, given that population variation in trait expression and dimorphism is well documented. Practitioners should also always go back to the original article and materials when scoring features and utilizing methods. That ensures that you are following the guidelines as the original author described, instead of someone else's interpretation of them. Finally, it is important for practitioners to be cognizant of the advantages and weaknesses to the various traits and methods and their potential implications in sex estimation. It will be interesting to see how cranial morphological sex estimation will continue to evolve with increased technology and a growing focus on quantifying traits through geometric morphometrics and computerized software.

References

Abdel Fatah, E. E., Shirley, N. R., Jantz, R. L., & Mahfouz, M. R. (2014). Improving sex estimation from crania using a novel three-dimensional quantitative method. *Journal of Forensic Sciences*, *59*(3), 590–600.

Acsádi, G., & Nemeskéri, J. (1970). *History of human life span and mortality*. Akadémiai Kiadó.

Baab, K. L., Freidline, S. E., Wang, S. L., & Hanson, T. (2010). Relationship of cranial robusticity to cranial form, geography and climate in *Homo sapiens*. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *141*(1), 97–115.

Bernal, V., Perez, S. I., & Gonzalez, P. N. (2006). Variation and causal factors of craniofacial robusticity in Patagonian hunter-gatherers from the late Holocene. *American Journal of Human Biology: The Official Journal of the Human Biology Association*, *18*(6), 748–765.

Boule, M., & Vallois, H. V. (1957). *Fossil men: A textbook of human palaeontology*. Dryden Press.

Broca, P. (1875). *Instructions craniologiques et craniométriques*. Masson.

Buikstra, J. E., & Ubelaker, D. H. (1994). Standardsfordata collectionfrom human skeletal remains. *Arkansas Archaeological Survey Research Series*, *44*.

Caple, J. M., Byrd, J. E., & Stephan, C. N. (2018). The utility of elliptical Fourier analysis for estimating ancestry and sex from lateral skull photographs. *Forensic Science International*.

Carrier, D. R., & Morgan, M. H. (2015). Protective buttressing of the hominin face. *Biological Reviews*, *90*(1), 330–346.

Daegling, D. J. (1993). Functional morphology of the human chin. *Evolutionary Anthropology: Issues, News, and Reviews*, *1*(5), 170–177.

Daubert (1993). Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579, 596.

Davies, D. M. (1972). *The influence of teeth, diet, and habits on the human face*. Butterworth-Heinemann.

Endo, B. (1970). Analysis of stresses around the orbit due to masseter and temporalis muscles respectively. *Journal of the Anthropological Society of Nippon*, *78*(4), 251–266.

Ferembach, D. (1980). Recommendations for age and sex diagnosis of skeletons. *Journal of Human Evolution*, *9*, 517–549.

Fiscella, G. N., & Smith, F. H. (2006). Ontogenetic study of the supraorbital region in modern humans: A longitudinal test of the spatial model. *Anthropologischer Anzeiger*, 147–160.

Garvin, H. M., & Klales, A. R. (2018). A validation study of the Langley et al. (2017) decision tree model for sex estimation. *Journal of Forensic Sciences*, *63*(4), 1243–1251.

Garvin, H. M., & Ruff, C. B. (2012). Sexual dimorphism in skeletal browridge and chin morphologies determined using a new quantitative method. *American Journal of Physical Anthropology*, *147*(4), 661–670.

Garvin, H. M., Sholts, S. B., & Mosca, L. A. (2014). Sexual dimorphism in human cranial trait scores: effects of population, age, and body size. *American Journal of Physical Anthropology*, *154*(2), 259–269.

Godinho, R. M., Spikins, P., & O'Higgins, P. (2018). Supraorbital morphology and social dynamics in human evolution. *Nature Ecology and Evolution*, *2*(6), 956.

Gonzalez, P. N., Bernal, V., & Perez, S. I. (2011). Analysis of sexual dimorphism of craniofacial traits using geometric morphometric techniques. *International Journal of Osteoarchaeology*, *21*(1), 82–91.

Guthrie, R. D. (1970). Evolution of human threat display organs. *Evolutionary Biology*, *4*(1), 257–302.

Hershkovitz, P. (1970). The decorative chin. *Bulletin of the American Museum of Natural History*, *41*, 6–10.

Horbaly, H. E., Kenyhercz, M. W., Hubbe, M., & Steadman, D. W. (2018). The influence of body size on the expression of sexually dimorphic morphological traits. *Journal of Forensic Sciences*.

Hylander, W. L., Johnson, K. R., & Picq, P. G. (1991). Masticatory-stress hypotheses and the supraorbital region of primates. *American Journal of Physical Anthropology*, *86*(1), 1–36.

Hylander, W. L., Picq, P. G., & Johnson, K. R. (1991). Function of the supraorbital region of primates. *Archives of Oral Biology*, *36*(4), 273–281.

Krantz, G. S. (1973). Cranial hair and brow ridges. *Mankind*, *9*(2), 109–111.

Krogman, W. M. (1955). The skeleton in forensic medicine. *Postgraduate Medicine*, *17*(2), A48–A62.

Kurten, B. (1979). The shadow of the brow. *Current Anthropology*, *20*(1), 229.

Langley, N. R., Dudzik, B., & Cloutier, A. (2018). A decision tree for nonmetric sex assessment from the skull. *Journal of Forensic Sciences*, *63*(1), 31–37.

Lewis, C. J., & Garvin, H. M. (2016). Reliability of the Walker cranial nonmetric method and implications for sex estimation. *Journal of Forensic Sciences*, *61*(3), 743–751.

Lieberman, D. E. (1995). Testing hypotheses about recent human evolution from skulls: Integrating morphology, function, development, and phylogeny. *Current Anthropology*, *36*(2), 159–197.

Moss, M. L., & Young, R. W. (1960). A functional approach to craniology. *American Journal of Physical Anthropology*, *18*(4), 281–292.

Murphy, R. E., & Garvin, H. M. (2018). A morphometric outline analysis of ancestry and sex differences in cranial shape. *Journal of Forensic Sciences*, *63*(4), 1001–1009.

National Academy of Sciences (NAS). (2009). *Strengthening forensic science in the United States: A path forward*. Washington, DC: National Academics Press.

Nikita, E., & Michopoulou, E. (2018). A quantitative approach for sex estimation based on cranial morphology. *American Journal of Physical Anthropology*, *165*(3), 507–517.

Oyen, O. J., Rice, R. W., & Samuel Cannon, M. (1979). Browridge structure and function in extant primates and Neanderthals. *American Journal of Physical Anthropology*, *51*(1), 83–95.

Penton-Voak, I. S., & Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evolution and Human Behavior*, *25*(4), 229–241.

Ravosa, M. J. (1988). Browridge development in Cercopithecidae: a test of two models. *American Journal of Physical Anthropology*, *76*(4), 535–555.

Riesenfeld, A. (1969). The adaptive mandible: an experimental study. *Cells Tissues Organs*, *72*(2), 246–262.

Rogers, T. L. (2005). Determining the sex of human remains through cranial morphology. *Journal of Forensic Science*, *50*(3), 1–8.

Russell, M. D., Brown, T., Garn, S. M., Giris, F., Turkel, S., İşcan, M. Y., et al. (1985). The supraorbital torus: "A Most Remarkable Peculiarity" [and comments and replies]. *Current Anthropology*, *26*(3), 337–360.

Stevenson, J. C., Mahoney, E. R., Walker, P. L., & Everson, P. M. (2009). prediction of sex based on five skull traits using decision analysis (CHAID). *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *139*(3), 434–441.

Walker, P. L. (1995). Problems of preservation and sexism in sexing: some lessons from historical collections for palaeodemographers. *Grave Reflections: Portraying the Past Through Cemetery Studies*, 31–47.

Walker, P. L. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *136*(1), 39–50.

Walls, M., Klales, A. R., Lesciotto, K., Gocha, T., & Garvin, H. M. (2018). Reliability and validity of Walker (2008) and Klales et al. (2012) methods. In *Poster presentation at the American Academy of Forensic Sciences 70th annual meeting, Seattle, WA*.

Walrath, D. E., Turner, P., & Bruzek, J. (2004). Reliability test of the visual assessment of cranial traits for sex determination. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *125*(2), 132–137.

Weidenreich, F. (1936). The Mandibles of Sinanthropus pekinensis: A comparative study Beijing. *Paleontologica Sinica, New Series D, 7*.

Williams, B. A., & Rogers, T. L. (2006). Evaluating the accuracy and precision of cranial morphological traits for sex determination. *Journal of Forensic Sciences*, *51*(4), 729–735.

CHAPTER 8

Analyses of the postcranial skeleton for sex estimation

Michala K. Stock
Department of Exercise Science, High Point University, High Point, NC, United States

Introduction and overview of sex estimation from the postcranium

The pelvis is widely regarded as the best skeletal region for estimating biological sex. However, there are many cases of incomplete or unassociated remains, as well as prevalent instances where the relatively fragile pubic bone has been damaged postmortem. In the absence of the pelvis, much research has favored both metric and morphological analyses of the skull—over postcranial bones besides those comprising the pelvis—to estimate sex. However, recent research has demonstrated that even simple, univariate analyses of single long bone measurements can yield similar discrimination accuracy between the sexes as multivariate analyses of the full suite of cranial measurements (Spradley & Jantz, 2011).

Furthermore, there may be cases where neither the pelvis nor the cranium is present for analysis, especially when remains are scattered or fragmented from various taphonomic agents, necessitating the creation of methods suitable for estimating sex from the postcranial skeleton. This issue has long been recognized in the field, with Dwight commenting on the variation in scapular maximum height (1894) as well as both humeral and femoral articular surface sizes between the sexes (1905), both displaying very little overlap in their ranges of variation between the sexes. The issue of potential subjectivity, bias, and effect of observer experience on using "anatomical appreciation"—what we now refer to as morphological or visual methods—was also commented upon in the early 20th century, instigating a call for greater use of metric comparison for estimating sex (Pearson, 1915).

In the intervening years, numerous studies have been published that focus on the metric assessment of various postcranial bones to inform sex estimation (Table 1). These methods include those focusing on a single measurement, such as the radial head diameter (Berrizbeitia, 1989), the humeral epicondylar breadth (Albanese et al., 2005), the femoral bicondylar breadth (Alunni-Perret et al., 2008), and the superoinferior femoral neck diameter (Frutos, 2003; Seidemann et al., 1998). Methods incorporating multiple measurements on a single element have also been proposed, such as those using the femur (Albanese et al., 2008; King et al., 1998; Mall et al., 2000; Purkait & Chandra, 2004),

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00008-0

Table 1 Measurements and populations used for postcranial metric sex estimation methods employing a single skeletal element.

Skeletal element	Measurement(s)	Citation	Population	Accuracy
Humerus	Epicondylar breadth	Albanese, Cardoso, and Saunders (2005)	Coimbra Collection	83%–96%
	22 measurements total, following variables informative for sex across populations: transverse and vertical humeral head diameters, mediolateral width of head + greater tubercle	France (1988)	Terry Collection; Sudanese Nubian archaeological sample (University of Colorado); Pecos Pueblo archaeological sample (Peabody Museum); Arikara archaeological sample	None listed
	Maximum length, maximum head diameter, maximum and minimum midshaft diameters, midshaft circumference, epicondylar breadth	Frutos (2005)	Modern Guatemalan forensic sample	76%–98%
	Maximum length, maximum head diameter, maximum and minimum midshaft diameter	Barnes and Wescott (2007)	Mississippian period (archaeological) sample from Missouri (University of Missouri—Columbia)	76%–84%
Radius	Maximum head diameter	Berrizbeitia (1989)	Terry Collection	92%–96%
Ulna	Olecranon-coronoid angle, inferior medial trochlear notch length and width	Purkait (2001)	Modern Indian sample (Bhopal Medicolegal Institute)	68%–96%
Femur	Bicondylar breadth	Alunni-Perret, Staccini, and Quatrehomme (2008)	Modern French donated sample (Medical School of Nice)	81%–97%
	Superoinferior neck diameter	Seidemann, Stojanowski, and Doran (1998)	Hamann-Todd Collection	87%–92%
		Frutos (2003)	Modern Guatemalan forensic sample	89%
	Maximum length, maximum head diameter, midshaft circumference, midshaft transverse and anteroposterior diameters, bicondylar breadth	King, Işcan, and Loth (1998)	Modern Thai sample (Chiang Mai University Hospital)	85%–94%

	Maximum length, transverse and vertical head diameter, head circumference, maximum midshaft diameter, bicondylar breadth	Mall, Graw, Gehring, and Hubig (2000)	Modern German sample (University of Cologne and University of Tűbingen)	67%–91%
	Maximum length, maximum head diameter, transverse and anteroposterior midshaft diameters, epicondylar breadth, midshaft circumference	Purkait and Chandra (2004)	Modern Indian sample (Bhopal Medicolegal Institute)	88%–99%
	Greater trochanter-fovea capits, fovea capitis-lesser trochanter, greater trochanter-lesser trochanter	Albanese, Eklics, and Tuck (2008)	Terry Collection	89%–95%
Tibia	Malleolar-condylar length, transverse and anteroposterior diameters and circumference at the nutrient foramen	İşcan and Miller-Shaivitz (1984)	Terry Collection	65%–83%
	Malleolar-condylar length, transverse and anteroposterior diameters and circumference at the nutrient foramen, distal and proximal epiphyseal breadths, minimum shaft circumference	İşcan, Yoshino, and Kato (1994)	Modern Japanese sample (Jikei Medical University)	80%–88%
	Tibial length, superior and inferior epiphyseal breadths, transverse and anteroposterior midshaft diameters, minimum circumference, circumference at the nutrient foramen	González-Reimers, Velasco-Vázquez, Arnay-De-La-Rosa, and Santolaria-Fernández (2000)	Archaeological (prehistoric) Gran Canaria population	94%–98%
	Full bone surfaces	Brzobohatá, Krajíček, Horák, and Velemínská (2016)	Medieval and early-20th-century, 21st-century samples from Czech Republic	76%–85%

Continued

Table 1 Measurements and populations used for postcranial metric sex estimation methods employing a single skeletal element—cont'd

Skeletal element	Measurement(s)	Citation	Population	Accuracy
Hyoid	Total breadth, body height and width, greater horn height and length	Komenda and Ĉerný (1990)	Modern Czech forensic sample	95%–96%
	31 measurements total, focus on total breadth and total (anteroposterior) length, and distal greater horn measurements	Miller, Walker, and O'Halloran (1998)	Modern US forensic sample (Ventura County, CA Medical Examiner's Office)	None listed
	13 measurements total, following variables used in discriminant function: height of the medial body, body thickness and breadth	Reesink, Van Immerseel, Brand, and Bruintjes (1999)	Modern Dutch sample (University of Leiden)	76%
	34 measurements total, following variables used in discriminant function: body breadth, greater horn height, length from narrowest point on greater horn to midpoint along long axis	Kim et al. (2006)	Modern Korean sample	88%
	Total breadth, anteroposterior length, body height and length, greater horn length, breadth at distal greater horns (fused), width and height of distal and proximal ends of greater horn	Kindschuh, Dupras, and Cowgill (2010)	Terry Collection	82%–85%
	Total breadth, anteroposterior length, greater horn length, slope of greater horns	Mukhopadhyay (2010)	Modern Indian forensic sample (Burdwan Medical College)	None listed
		Pollard et al. (2011)	Modern French forensic sample (Aix-Marseille University)	None listed
	Total breadth, anteroposterior length, greater horn minimum diameter	D'Souza, Kiran, and Harish (2013)	Modern Indian forensic sample (including juvenile and adolescent decedents)	None listed
	Total breadth, anteroposterior length, body height, length and thickness, greater horn height and length, 23 landmarks covering bone (for shape)	Urbanová, Hejna, Zátopková, and Šafr (2013)	Modern Czech forensic sample	82%–96%

C2	Maximum sagittal length, maximum height and sagittal and transverse breadths of the dens, vertebral foramen anteroposterior length, superior facet sagittal and transverse diameters, maximum breadth across the superior facets	Wescott (2000)	Hamann-Todd Collection; Terry Collection	78%–90%
Scapula	Maximum height	Dwight (1894)	Harvard Medical School	None listed
		Dabbs (2009)	Hamann-Todd Collection	92%–100%
	Maximum height and breadth, maximum length of spine, glenoid height, lateral curvature, lateral border thickness	Dabbs and Moore-Jansen (2010)	Hamann-Todd Collection	89%–96%
	Maximum height and breadth, glenoid length and breadth, coracoid length, acromion length, coracoid-acromion distance	Di Vella, Campobasso, Dragone, and Introna (1994)	Modern Italian forensic sample (University of Bari)	68%–91%
	Area of glenoid cavity	Prescher and Klümpen (1995)	Modern German sample	36%–60%
	Maximum height and breadth, glenoid length and breadth	Özer, Katayama, Sahgir, and Güleç (2006)	Medieval Anatolian sample	82%–95%
	21 landmarks around the scapular body (2D)	Scholtz, Steyn, and Pretorius (2010)	Modern South African sample (Pretoria Bone Collection)	91%–95%
	Maximum height and breadth	Ali et al. (2018)	Modern US forensic sample (Office of Chief Medical Examiner for the state of Maryland)	94%

Continued

Table 1 Measurements and populations used for postcranial metric sex estimation methods employing a single skeletal element—cont'd

Skeletal element	Measurement(s)	Citation	Population	Accuracy
Ribs	Supero-inferior height and antero-posterior thickness of fourth rib sternal end	Wiredu, Kumoji, Seshadri, and Biritwum (1999)	Modern West African forensic sample (Korle Bu Teaching Hospital)	74%–80%
Patella	Maximum height and width, thickness, maximum heights and widths of the medial and lateral articular surfaces	Introna, Di Vella, and Campobasso (1998)	Modern Italian forensic sample (University of Bari)	71%–83%
	Maximum breadth, height, and thickness of bone, maximum height of articular facet, maximum breadths of medial and lateral articular facets	Bidmos, Steinberg, and Kuykendall (2005) Dayal and Bidmos (2005)	Dart Collection (South African Whites) Dart Collection (South African Blacks)	77%–85% 77%–85%
	25 variables extracted via automated means from CT scans	Mahfouz et al. (2007)	Bass Donated Skeletal Collection	83%–96%
Calcaneus	Maximum length, breadth, and height, load-arm length, dorsal articular facet length and breadth, body height, cuboid facet height, middle breadth	Bidmos and Asala (2004)	Dart Collection	64%–86%
	Maximum length, load-arm length and width, posterior circumference	DiMichele and Spradley (2012)	Bass Donated Skeletal Collection	84%–88%
Talus	Total length, width, and height, trochlear length and breadth, head-neck length, height of the head, posterior articular surface length and breadth	Bidmos and Dayal (2003)	Dart Collection	80%–88%

the tibia (Brzobohatá et al., 2016; González-Reimers et al., 2000; İşcan & Miller-Shaivitz, 1984; İşcan et al., 1994), the humerus (Barnes & Wescott, 2007; France, 1988; Frutos, 2005), the ulna (Purkait, 2001), the hyoid (D'Souza et al., 2013; Kim et al., 2006; Kindschuh et al., 2010; Komenda & Ĉerný, 1990; Miller et al., 1998; Mukhopadhyay, 2010; Pollard et al., 2011; Reesink et al., 1999; Urbanová et al., 2013), the second cervical vertebra (C2; Wescott, 2000), the scapula (Ali et al., 2018; Dabbs, 2009; Dabbs & Moore-Jansen, 2010; Di Vella et al., 1994; Dwight, 1894; Özer et al., 2006; Prescher & Klümpen, 1995; Scholtz et al., 2010), the sternal rib end (Wiredu et al., 1999), the patella (Bidmos et al., 2005; Dayal & Bidmos, 2005; Introna et al., 1998; Mahfouz et al., 2007), the calcaneus (Bidmos & Asala, 2004; DiMichele & Spradley, 2012), and the talus (Bidmos & Dayal, 2003). And finally, multivariate analyses using multiple measurements from multiple elements have also been put forth (Table 2), such as incorporating several measurements from five major long bones of the limbs (excluding the fibula; Safont et al., 2000; Sakaue, 2004; Steel, 1962; Wrobel et al., 2002); using articulating elements such as the femur and os coxa (Albanese, 2003), the femur and tibia (Steyn & İşcan, 1997), the long bones of the arm (Albanese, 2013; Charisi et al., 2011; Holman & Bennett, 1991; Mall et al., 2001), the shoulder girdle (Frutos, 2002; Murphy, 2002; Van Dongen, 1963), the metacarpals (Barrio et al., 2006), the metatarsals (Robling & Ubelaker, 1997), or the talus and calcaneus (Gualdi-Russo, 2007; Steele, 1976); as well as combining measurements from homologous elements like the humerus and femur (Boldsen et al., 2015), or the manual and pedal long bones (Case & Ross, 2007). As may be appreciated from the above—by no means exhaustive—list of examples, much work has focused on the metric assessment of sexual dimorphism in the postcranial skeleton over the past century. Relatively less research has been performed utilizing morphological traits for estimating sex, and these methods will be reviewed in greater depth below (Table 3).

Rogers et al. (2000) examined the inferior surface of the medial clavicular shaft, where the costoclavicular (or rhomboid) ligament connects the clavicle and the first rib, and found sexually dimorphic expression of the rhomboid fossa. This study was conducted on pairs of clavicles from 113 females and 231 males (total $N=344$) from the William F. McCormick Collection, and then tested on a smaller subsample of the William M. Bass Skeletal Collection. The presence of a deep fossa with potential trabecular bone exposure, rather than a slight impression or a raised surface, suggests that the individual is male, with 92.2% posterior probability on the left side and 81.7% on the right. Following the publication of this method, numerous studies using both dry bone and radiographic visual inspection have corroborated the significantly higher prevalence of rhomboid fossae in males among several diverse populations (Ishwarkumar et al., 2016; Kaewma et al., 2016; Koudela et al., 2015; Prado et al., 2009; Sehrawat & Pathak, 2016; Singh & Singh, 2009; Vani et al., 2018).

Table 2 Measurements and populations used for postcranial metric sex estimation methods employing multiple skeletal elements.

Skeletal elements	Measurements	Citation	Population	Accuracy
Five major limb long bones (excluding fibula)	Maximum lengths of the humerus, radius, ulna, and femur; vertical humeral head diameter, horizontal femoral head diameter, humeral and femoral epicondylar breadths, radial and ulnar mediolateral distal breadths, radial tuberosity diameter, ulnar coronoid height, tibial antero-posterior diameter at the nutrient foramen	Steel (1962)	St. Bride's Church (historical) sample	None listed
	Minimum shaft circumferences of the humerus, radius, ulna, and tibia; circumferences at the radial tuberosity, the femoral midshaft and the subtrochanter	Safont, Malgosa, and Subirà (2000)	Roman-period and modern Spanish samples	80%–92%
	Circumferences at the femoral and tibial midshafts; minimum circumferences of the tibial, humeral, radial, and ulnar shafts; maximum diameters of the femoral and humeral heads; maximum breadth at the anteroposterior femoral shaft, humeral midshaft, the deltoid tuberosity, and the radial tuberosity; anteroposterior breadths at the femoral subtrochanter and midshaft, tibial midshaft and nutrient foramen	Wrobel, Danforth, and Armstrong (2002)	Archaeological (prehistoric) Mayan sample	77%–100%
	47 measurements total, following variables "useful" for sex estimation: distal humeral articular breadth, radial head sagittal diameter, ulnar midshaft area, femoral bicondylar breadth, tibial proximal epiphyseal breadth	Sakaue (2004)	Modern Japanese sample (University of Tokyo and Chiba University)	91%–95%
Femur and os coxa	Os coxa: pelvic height, iliac breadth, pubis length, ischium length; femur: maximum length, maximum head diameter, epicondylar breadth	Albanese (2003)	Terry Collection; Coimbra Collection	90%–98%

Femur and tibia	Femur: maximum length, maximum head diameter, midshaft circumference, transverse and anteroposterior diameters, epicondylar breadth; tibia: physiological length, epicondylar breadth, circumference, transverse and anteroposterior diameters at nutrient foramen, minimum distal shaft circumference, distal epiphyseal breadth	Steyn and İşcan (1997)	Dart Collection; modern South African forensic sample (University of Pretoria)	86%–98%
Humerus, radius, and ulna	Maximum humeral, radial, and ulnar lengths, distal radial and ulnar mediolateral breadths	Holman and Bennett (1991)	Terry Collection	84%–96%
	Maximum humeral, radial, and ulnar lengths, humeral vertical head diameter and epicondylar breadth, maximum radial and ulnar distal breadths, maximum radial head diameter, maximum proximal ulnar breadth	Mall et al. (2001)	Modern German samples (Universities Cologne & Munich)	72%–90%
		Charisi, Eliopoulos, Vanna, Koilias, and Manolis (2011)	Modern Greek sample (University of Athens)	90%–95%
Bones of the forelimb	Clavicle: maximum length, midshaft superior-inferior diameter; humerus: vertical head diameter, epicondylar breadth; radius: maximum length, maximum head diameter, midshaft anteroposterior diameter; ulna: maximum and perpendicular diameter at maximal crest development	Albanese (2013)	Terry Collection; Coimbra Collection	87%–97%
Bones of the shoulder girdle	Clavicle: maximum length, minimum shaft circumference, sternal and acromial breadths; scapula: maximum length and breadth, spine length, coraco-acromial breadth, glenoid fossa height and breadth, length-breadth and axillo-spinal angles; humerus: maximum and physiological lengths, circumferences at the head, surgical neck, and minimum shaft, vertical and transverse head diameters, maximum proximal and distal epiphyseal breadths, maximum and minimum midshaft diameters, angle of torsion	Van Dongen (1963)	Prehistoric Indigenous Australian sample (South Australian Museum)	None listed

Continued

Table 2 Measurements and populations used for postcranial metric sex estimation methods employing multiple skeletal elements—cont'd

Skeletal elements	Measurements	Citation	Population	Accuracy
	Clavicle: maximum length and midshaft circumference; scapula: glenoid fossa height and width	Frutos (2002)	Modern Guatemalan forensic sample	85%–94%
	Clavicle: diameters at acromial and sternal ends; scapula: glenoid fossa height and width	Murphy (2002)	Prehistoric Polynesian sample (Otago School of Medical Sciences)	97%
Metacarpals	For all five metacarpals: maximum lengths, epicondylar breadths, anteroposterior and mediolateral diameters of the proximal epiphyses, midshafts, and distal epiphyses	Barrio, Trancho, and Sanchez (2006)	Modern Spanish sample (Complutense University of Madrid)	81%–91%
Metatarsals	For all five metatarsals: maximum lengths, midshaft diameters, superoinferior and mediolateral diameters of the heads and bases	Robling and Ubelaker (1997)	Terry Collection	83%–100%
Talus and calcaneus	Talus and calcaneus: length, width, and height (right and left sides)	Gualdi-Russo (2007)	Modern Italian sample (University of Bologna)	87%–95%
	Talus: maximum length, width, and height, maximum trochlear length and width; calcaneus: maximum length, minimum width, body height, load-arm length and width	Steele (1976)	Terry Collection	79%–89%
Humerus and femur	Humerus: vertical head diameter and epicondylar breadth; femur: maximum head diameter	Boldsen, Milner, and Boldsen (2015)	Bass Donated Skeletal Collection; Modern US forensic sample (Mercyhurst University)	72%–88%
Manual and pedal long bones	Hand: metacarpal and all manual phalangeal lengths; foot: metatarsal and first ray phalangeal lengths	Case and Ross (2007)	Terry Collection	74%–85%

Table 3 Skeletal regions and traits used for postcranial morphological sex estimation.

Skeletal region	Trait(s)	Citation	Population	Accuracy	Validation studies
Medial clavicle	Rhomboid fossa	Rogers, Flournoy, and McCormick (2000)	William F. McCormick Skeletal Collection	81%–92%	Prado et al. (2009), Singh and Singh (2009), Koudela, Koudelová, and Zeman (2015), Ishwarkumar, Pillay, Haffajee, and Rennie (2016), Kaewma, Sampannang, Tuamsuk, Kanpittaya, and Iamsaard (2016), Sehrawat and Pathak (2016), and Vani, Malsawmzuali, Anbalagan, and Rajasekar (2018)
Distal humerus	Constriction of the trochlea Degree of trochlear asymmetry Olecranon fossa shape Angle of the medial epicondyle	Rogers (1999)	University of Toronto Grant Skeletal Collection	92%	Falys, Schutkowski, and Weston (2005) and Rogers (2009) (adolescents); Vance, Steyn, and L'Abbe (2011)

The other area of the postcranial skeleton that has received attention for the morphological assessment of sex is the distal humerus. Rogers (1999) created a tripartite scoring system for four traits on the distal humerus: the constriction of the trochlea, degree of asymmetry of the trochlea, the shape of the olecranon fossa, and the angle of the medial epicondyle. Compared to males, females tend to display more constricted and symmetrical trochleae, deeper and oval-shaped olecranon fossae, and angled medial epicondyles. When all four traits were employed, this study demonstrated a 92% accuracy rate (Rogers, 1999). A blind test of Rogers' method found similarly high accuracy for discriminating sex based on visual examination of these same traits in the distal humerus for an archaeological population (Falys et al., 2005). This method was also tested on a sample of modern black and white South Africans using three of the four original traits (trochlear constriction was excluded in this study), and the authors found accuracy rates of 74% and 77% for males and females, respectively (Vance et al., 2011). These findings suggest that further investigation into the differences in sexual dimorphism in these traits across populations is appropriate. Additionally, because the distal humeral epiphysis begins fusing to the epiphysis relatively early in adolescence, the Rogers method may also be applicable to skeletally immature individuals (Rogers, 2009).

Some limited work has been done assessing morphological sex-based variation in the scapula (Bainbridge & Tarazaga, 1956; Van Dongen, 1963) and the hyoid (D'Souza et al., 2013), neither of which yielded results useful for estimating sex. Bainbridge and Tarazaga reviewed a suite of morphological characteristics from areas across the scapula, but found "no really useful criteria of sex for the scapula" (1956, p. 110; neither did Van Dongen, 1963). D'Souza et al. (2013) assessed the shape of the adult hyoid bone as either "U-" or "V-shaped," but the distribution of these shapes between the sexes did not approach statistical significance.

Sex-based differences in skeletal robusticity and rugosity

Much of the research regarding methods for assessing sex from the postcranial bones relies on the observation that in humans (and among most primates) within a given population, males tend to be larger than females, including both taller statures and greater body masses. As discussed above, this has manifested in the practice of sex estimation with a focus on distinguishing metric cut-offs between relatively larger (male) and relatively smaller (female) joint surfaces (e.g., Stewart, 1979). More recently, cross-sectional shape and the degree of robusticity in the shafts of postcranial elements has received more attention (Carlson, Grine, & Pearson, 2007; Pomeroy & Zakrzewski, 2009; Ruff, 1987). These studies focus on archaeological populations, where sex-based differences in activity patterns, such as degree of mobility and tool use, and the associated differences in biomechanical environments are postulated to have created the sexual dimorphism in long bone cross-sectional shape displayed by many hunter-gatherer populations (Carlson et al., 2007; Pomeroy & Zakrzewski, 2009; Ruff, 1987). However, for this to be useful for estimating sex, two a priori assumptions must be met: (1) that there is a discernable level of sexual dimorphism in diaphyseal shape and/or size, and (2) that the patterns of these sex-based differences in the given population are known via previous skeletal studies or correctly inferred from the archaeological or ethnographic record. Comparing a prehistoric, archaeological population to a modern, industrialized sample revealed a marked reduction in the degree of sexual dimorphism in diaphyseal shape (Ruff, 1987). Furthermore, evidence of sex-based division of labor may not be available for all bioarchaeological scenarios, and is virtually impossible to infer in a forensic context. These considerations suggest substantial limitations for application to forensically significant cases.

Discussions of sex-based differences in both robusticity, as above, and rugosity, see below, are predicated on the idea that mechanical loading—whether by gravitational or muscular forces (for a review, see Judex & Carlson, 2009)—influences bones' morphology (e.g., Frost, 2001; Ruff, Holt, & Trinkaus, 2006). Bone functional adaptation is an important consideration since bones, of course, do not exist in a vacuum; and as a living tissue, bone can change in response to mechanical stimuli. For these reasons, the mechanical effect of the muscles on bones, seen at muscle attachment sites, has also been an area of recent interest for investigating sexual dimorphism in the postcranial skeleton.

Like diaphyseal robusticity, enthesis morphology, or the rugosity of bone at muscle attachment sites, has been assumed to be useful for assessing activity patterns from past populations. However, recent empirical testing of this assumption in animal models demonstrates no correlation between enthesis morphology on muscle mass or activity patterns (e.g., Zumwalt, 2006). If enthesis rugosity does not correspond to differential loading regimes, then something other than activity patterns must create variation at these sites.

Rather than sex, age was found to be the best predictor of muscle marking rugosity when muscle attachment sites on the long bones of the upper limb (Weiss, 2003) as well as the lower limb (Weiss, 2004) are aggregated. However, these studies also demonstrated collinearity in that individuals who had the greatest expression of muscle markings tended to concomitantly also be male and have larger, more robust long bones (Weiss, 2003, 2004), leading this author to call for controlling both element articular size and age when investigating sex-based differences in muscle markings (Weiss, 2007). Yet, even when these variables (articular size and age) are accounted for, it appears that sex-based differences in muscle marker expression may be confounded by the effect of body mass, as this is also a sexually dimorphic trait in humans (Weiss, Corona, & Schultz, 2012).

Methodological issues aside from collinearity or confounding variables have also been noted, mainly regarding lack of consensus on the optimal ways to measure and/or score entheses, resulting in a large number of available methods (Foster, Buckley, & Tayles, 2014). In sum, the complexity of the biological processes and the multitude of factors playing into enthesis formation suggest that analyses of muscle attachment site rugosity for inferring activity pattern and/or estimating sex from skeletal material should be undertaken with caution, if performed at all (Foster et al., 2014).

Morphological postcranial sex estimation in practice

There are relatively few morphological methods for sex estimation from the postcranial skeleton; and those that exist should be used in conjunction with other types of analyses or compared against the sex estimation results from other skeletal elements. Therefore, the methods described in this chapter are not routinely highlighted in forensic anthropological casework in the author's experience. However, because instances encountered by forensic anthropologists and bioarchaeologists may involve fragmentary and/or incomplete skeletal remains, the creation, validation, and use of such methods continues to be of value. Despite the caveats noted above, observations of general degree of robusticity of postcranial elements and rugosity of muscle attachments are routinely noted and may tentatively contribute to the preponderance of evidence considered—especially in cases of commingling—when estimating biological sex for a given individual.

Conclusion

The pelvis and skull continue to be the preferred skeletal elements for estimating sex; however, multiple methods have been created and used to estimate sex from the postcranial skeleton as well. These include a multitude of metric methods, predominantly focusing on comparing measurements of joint surface, as well as several morphological methods. Visual inspection of both the rhomboid fossa of the clavicle (Rogers et al., 2000) and a suite of traits on the distal humerus (Rogers, 1999) demonstrates high accuracy rates, including in validation studies on temporally and geographically diverse populations. As the distal humeral epiphysis fuses relatively early in adolescence, the latter method may even be applicable for assessing sex in adolescent decedents (Rogers, 2009). Alternate methods derived from archaeological populations rely on sex-based division of labor and gendered activity patterns to create sex-based differences in the robusticity and rugosity of postcranial skeletal elements, which is problematic from both theoretical and methodological perspectives, and these traits should be employed with caution. While the literature on morphological postcranial sex estimation is relatively limited at present, the need for methods that are easy to implement on fragmentary/incomplete remains indicates that this is an area for fruitful future research.

References

Albanese, J. (2003). A metric method for sex determination using the hipbone and femur. *Journal of Forensic Sciences*, *48*, 1–11.

Albanese, J. (2013). A method for estimating sex using the clavicle, humerus, radius, and ulna. *Journal of Forensic Sciences*, *58*, 1413–1419.

Albanese, J., Cardoso, H. F., & Saunders, S. R. (2005). Universal methodology for developing univariate sample-specific sex determination methods: An example using the epicondylar breadth of the humerus. *Journal of Archaeological Research*, *32*, 143–152.

Albanese, J., Eklics, G., & Tuck, A. (2008). A metric method for sex determination using the proximal femur and fragmentary hipbone. *Journal of Forensic Sciences*, *53*, 1283–1288.

Ali, Z., Cox, C., Stock, M. K., Zandee van Rilland, E. E., Rubio, A., & Fowler, D. R. (2018). Estimating sex using metric analysis of the scapula by postmortem computed tomography. *Journal of Forensic Sciences*. https://doi.org/10.1111/1556-4029.13751.

Alunni-Perret, V., Staccini, P., & Quatrehomme, G. (2008). Sex determination from the distal part of the femur in a French contemporary population. *Forensic Science International*, *175*, 113–117.

Bainbridge, D., & Tarazaga, S. G. (1956). A study of sex differences in the scapula. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland*, *86*, 109–134.

Barnes, J., & Wescott, D. J. (2007). Sex determination of Mississippian skeletal remains from human measurements. *The Missouri Archaeologist*, *68*, 133–137.

Barrio, P. A., Trancho, G. J., & Sanchez, J. A. (2006). Metacarpal sexual determination in a Spanish population. *Journal of Forensic Sciences*, *51*, 990–995.

Berrizbeitia, E. L. (1989). Sex determination with the head of the radius. *Journal of Forensic Sciences*, *34*, 1206–1213.

Bidmos, M., & Dayal, M. (2003). Further evidence to show population specificity of discriminant function equations for sex determination using the talus of South African blacks. *Journal of Forensic Sciences*, *49*, 1165–1170.

Bidmos, M. A., & Asala, S. A. (2004). Sexual dimorphism of the calcaneus of South African blacks. *Journal of Forensic Sciences*, *49*, 446–450.

Bidmos, M. A., Steinberg, N., & Kuykendall, K. L. (2005). Patella measurements of South African whites as sex assessors. *Homo*, *56*, 69–74.

Boldsen, J. L., Milner, G. R., & Boldsen, S. K. (2015). Sex estimation from modern American humeri and femora, accounting for sample variance structure. *American Journal of Physical Anthropology*, *158*, 745–750.

Brzobohatá, H., Krajíček, V., Horák, Z., & Velemínská, J. (2016). Sexual dimorphism of the human tibia through time: Insights into shape variation using a surface-based approach. *PloS ONE*, *11*, e0166461.

Carlson, K. J., Grine, F. E., & Pearson, O. M. (2007). Robusticity and sexual dimorphism in the postcranium of modern hunter-gatherers from Australia. *American Journal of Physical Anthropology*, *134*, 9–23.

Case, D. T., & Ross, A. H. (2007). Sex determination from hand and foot bone lengths. *Journal of Forensic Sciences*, *52*, 264–270.

Charisi, D., Eliopoulos, C., Vanna, V., Koilias, C. G., & Manolis, S. K. (2011). Sexual dimorphism of the arm bones in a modern Greek population. *Journal of Forensic Sciences*, *56*, 10–18.

D'Souza, D. H., Kiran, J., & Harish, S. S. (2013). Determination of sex by shape and size of hyoid bone. *Journal of Indian Academy of Forensic Medicine*, *35*, 145–147.

Dabbs, G. R. (2009). Is Dwight right? Can the maximum height of the scapula be used for accurate sex estimation? *Journal of Forensic Sciences*, *54*, 529–530.

Dabbs, G. R., & Moore-Jansen, P. H. (2010). A method for estimating sex using metric analysis of the scapula. *Journal of Forensic Sciences*, *55*, 149–152.

Dayal, M., & Bidmos, M. (2005). Discriminating sex in South African blacks using patella dimensions. *Journal of Forensic Sciences*, *50*, 1294–1297.

Di Vella, G., Campobasso, C. P., Dragone, M., & Introna, F. (1994). Skeletal sex determination by scapular measurements. *Bollettino - Societa Italiana di Biologia Sperimentale*, *70*, 299–305.

DiMichele, D. L., & Spradley, M. K. (2012). Sex estimation in a modern American osteological sample using a discriminant function analysis from the calcaneus. *Forensic Science International*, *221*, 152.e151–152.e155.

Dwight, T. (1894). The range and significance of variation in the human skeleton: The Shattuck lecture for 1894. *The Boston Medical and Surgical Journal*, *131*, 73–76.

Dwight, T. (1905). The size of the articular surfaces of the long bones as characteristic of sex. *The American Journal of Anatomy*, *4*, 19–31.

Falys, C. G., Schutkowski, H., & Weston, D. A. (2005). The distal humerus-a blind test of Rogers' sexing technique using a documented skeletal collection. *Journal of Forensic Sciences*, *50*, 171–175.

Foster, A., Buckley, H., & Tayles, N. (2014). Using enthesis robusticity to infer activity in the past: A review. *Journal of Archaeological Method and Theory*, *21*, 511–533.

France, D. L. (1988). Osteometry at muscle origin and insertion in sex determination. *American Journal of Physical Anthropology*, *76*, 515–526.

Frost, H. M. (2001). From Wolff's law to the Utah paradigm: Insights about bone physiology and its clinical applications. *The Anatomical Record*, *262*, 398–419.

Frutos, L. R. (2002). Determination of sex from the clavicle and scapula in a Guatemalan contemporary rural indigenous population. *The American Journal of Forensic Medicine and Pathology*, *23*, 284–288.

Frutos, L. R. (2003). Brief communication: Sex determination accuracy of the minimum supero-inferior femoral neck diameter in a contemporary rural Guatemalan population. *American Journal of Physical Anthropology*, *122*, 123–126.

Frutos, L. R. (2005). Metric determination of sex from the humerus in a Guatemalan forensic sample. *Forensic Science International*, *147*, 153–157.

González-Reimers, E., Velasco-Vázquez, J., Arnay-De-La-Rosa, M., & Santolaria-Fernández, F. (2000). Sex determination by discriminant function analysis of the right tibia in the prehispanic population of the Canary Islands. *Forensic Science International*, *108*, 165–172.

Gualdi-Russo, E. (2007). Sex determination from the talus and calcaneus measurements. *Forensic Science International*, *171*, 151–156.

Holman, D. J., & Bennett, K. A. (1991). Determination of sex from arm bone measurements. *American Journal of Physical Anthropology*, *84*, 421–426.

Introna, F., Di Vella, G., & Campobasso, C. P. (1998). Sex determination by discriminant analysis of patella measurements. *Forensic Science International*, *95*, 39–45.

İşcan, M. Y., & Miller-Shaivitz, P. (1984). Determination of sex from the tibia. *American Journal of Physical Anthropology*, *64*, 53–57.

İşcan, M. Y., Yoshino, M., & Kato, S. (1994). Sex determination from the tibia: Standards for contemporary Japan. *Journal of Forensic Sciences*, *39*, 785–792.

Ishwarkumar, S., Pillay, P., Haffajee, M., & Rennie, C. (2016). Sex determination using morphometric and morphological dimensions of the clavicle within the KwaZulu-Natal population. *International Journal of Morphology*, *34*, 244–251.

Judex, S., & Carlson, K. J. (2009). Is bone's response to mechanical signals dominated by gravitational loading? *Medicine and Science in Sports and Exercise*, *41*, 2037–2043.

Kaewma, A., Sampannang, A., Tuamsuk, P., Kanpittaya, J., & Iamsaard, S. (2016). Incidence of clavicular rhomboid fossa in northeastern Thais: An anthropological study. *Anatomy Research International*, *2016*, 1–4.

Kim, D. I., Lee, U. Y., Park, D. K., Kim, Y. S., Han, K. H., Kim, K. H., & Han, S. H. (2006). Morphometrics of the hyoid bone for human sex determination from digital photographs. *Journal of Forensic Sciences*, *51*, 979–984.

Kindschuh, S. C., Dupras, T. L., & Cowgill, L. W. (2010). Determination of sex from the hyoid bone. *American Journal of Physical Anthropology*, *143*, 279–284.

King, C. A., Işcan, M. Y., & Loth, S. R. (1998). Metric and comparative analysis of sexual dimorphism in the Thai femur. *Journal of Forensic Sciences*, *43*, 954–958.

Komenda, S., & Ĉerný, M. (1990). Sex determination from the hyoid bone by means of discriminant analysis. *Acta Universitatis Palackianae Olomucensis Facultatis Medicae*, *125*, 37–51.

Koudela, K., Koudelová, J., & Zeman, P. (2015). Concave impressio ligamenti costoclavicularis ("rhomboid fossa") and its prevalence and relevance to clinical practice. *Surgical and Radiologic Anatomy*, *37*, 239–245.

Mahfouz, M., Badawi, A., Merkl, B., Fatah, E. E. A., Pritchard, E., Kesler, K., … Jantz, L. (2007). Patella sex determination by 3D statistical shape models and nonlinear classifiers. *Forensic Science International*, *173*, 161–170.

Mall, G., Graw, M., Gehring, K. D., & Hubig, M. (2000). Determination of sex from femora. *Forensic Science International*, *113*, 315–321.

Mall, G., Hubig, M., Büttner, A., Kuznik, J., Penning, R., & Graw, M. (2001). Sex determination and estimation of stature from the long bones of the arm. *Forensic Science International*, *117*, 23–30.

Miller, K. W., Walker, P. L., & O'Halloran, R. L. (1998). Age and sex-related variation in hyoid bone morphology. *Journal of Forensic Sciences*, *43*, 1138–1143.

Mukhopadhyay, P. P. (2010). Morphometric features and sexual dimorphism of adult hyoid bone: A population specific study with forensic implications. *Journal of Forensic and Legal Medicine*, *17*, 321–324.

Murphy, A. M. C. (2002). Articular surfaces of the pectoral girdle: Sex assessment of prehistoric New Zealand Polynesian skeletal remains. *Forensic Science International*, *125*, 134–136.

Özer, I., Katayama, K., Sahgir, M., & Güleç, E. (2006). Sex determination using the scapula in medieval skeletons from East Anatolia. *Collegium Antropologicum*, *30*, 415–419.

Pearson, K. (1915). On the problem of sexing osteometric material. *Biometrika*, *10*, 479–487.

Pollard, J., Piercecchi-Marti, M. D., Thollon, L., Bartoli, C., Adalian, P., Bécart-Robert, A., … Léonetti, G. (2011). Mechanisms of hyoid bone fracture after modelling: Evaluation of anthropological criteria defining two relevant models. *Forensic Science International*, *212*, 274.e1–274.e5.

Pomeroy, E., & Zakrzewski, S. R. (2009). Sexual dimorphism in diaphyseal cross-sectional shape in the medieval Muslim population of Écija, Spain, and Anglo-Saxon Great Chesterford, UK. *International Journal of Osteoarchaeology*, *19*, 50–65.

Prado, F., de Mello Santos, L., Caria, P., Kawaguchi, J., Preza, A. d. O., Daruge, J. E., … Daruge, E. (2009). Incidence of clavicular rhomboid fossa (impression for costoclavicular ligament) in the Brazilian population: Forensic application. *The Journal of Forensic Odonto-Stomatology*, *27*, 12–16.

Prescher, A., & Klümpen, T. (1995). Does the area of the glenoid cavity of the scapula show sexual dimorphism? *Journal of Anatomy*, *186*, 223.

Purkait, R. (2001). Measurements of ulna—A new method for determination of sex. *Journal of Forensic Sciences*, *46*, 924–927.

Purkait, R., & Chandra, H. (2004). A study of sexual variation in Indian femur. *Forensic Science International*, *146*, 25–33.

Reesink, E. M., Van Immerseel, A. A. H., Brand, R., & Bruintjes, T. D. (1999). Sexual dimorphism of the hyoid bone? *International Journal of Osteoarchaeology*, *9*, 357–360.

Robling, A. G., & Ubelaker, D. H. (1997). Sex estimation from the metatarsals. *Journal of Forensic Sciences*, *42*, 1062–1069.

Rogers, N., Flournoy, L., & McCormick, W. (2000). The rhomboid fosssa of the clavicle as a sex and age estimator. *Journal of Forensic Sciences*, *45*, 61–67.

Rogers, T. L. (1999). A visual method of determining the sex of skeletal remains using the distal humerus. *Journal of Forest Science*, *51*, 57–60.

Rogers, T. L. (2009). Sex determination of adolescent skeletons using the distal humerus. *American Journal of Physical Anthropology*, *140*, 143–148.

Ruff, C. (1987). Sexual dimorphism in human lower limb bone structure: Relationship to subsistence strategy and sexual division of labor. *Journal of Human Evolution*, *16*, 391–416.

Ruff, C., Holt, B., & Trinkaus, E. (2006). Who's afraid of the big bad Wolff?: "Wolff's law" and bone functional adaptation. *American Journal of Physical Anthropology*, *129*, 484–498.

Safont, S., Malgosa, A., & Subirà, M. E. (2000). Sex assessment on the basis of long bone circumference. *American Journal of Physical Anthropology*, *113*, 317–328.

Sakaue, K. (2004). Sexual determination of long bones in recent Japanese. *Anthropological Science*, *112*, 75–81.

Scholtz, Y., Steyn, M., & Pretorius, E. (2010). A geometric morphometric study into the sexual dimorphism of the human scapula. *Homo*, *61*, 253–270.

Sehrawat, J. S., & Pathak, R. (2016). Variability in anatomical features of human clavicle: Its forensic anthropological and clinical significance. *Translational Research in Anatomy*, *3*, 5–14.

Seidemann, R., Stojanowski, C., & Doran, G. (1998). The use of the supero-inferior femoral neck diameter as a sex estimator. *American Journal of Physical Anthropology*, *107*, 305–313.

Singh, J., & Singh, D. (2009). Rhomboid fossa of the clavicle: Is it a reliable estimator of gender and age of Northwest Indian subjects of Chandigarh Zone. *Journal of Punjab Academy of Forensic Medicine & Toxicology*, *9*, 58–65.

Spradley, M. K., & Jantz, R. L. (2011). Sex estimation in forensic anthropology: Skull versus postcranial elements. *Journal of Forensic Sciences*, *56*, 289–296. https://doi.org/10.1111/j.1556-4029.2010.01635.x.

Steel, F. L. D. (1962). The sexing of long bones, with reference to the St Bride's series of identified skeletons. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland*, *92*, 212–222.

Steele, D. (1976). The estimation of sex on the basis of the talus and calcaneus. *American Journal of Physical Anthropology*, *45*, 581–588.

Stewart, T. (1979). *Essentials of forensic anthropology*. Springfield, IL: Charles C Thomas.

Steyn, M., & İşcan, M. (1997). Sex determination from the femur and tibia in South African whites. *Forensic Science International*, *90*, 111–119.

Urbanová, P., Hejna, P., Zátopková, L., & Šafr, M. (2013). What is the appropriate approach in sex determination of hyoid bones? *Journal of Forensic and Legal Medicine*, *20*, 996–1003.

Van Dongen, R. (1963). The shoulder girdle and humerus of the Australian Aborigine. *American Journal of Physical Anthropology*, *21*, 469–488.

Vance, V., Steyn, M., & L'Abbe, E. (2011). Nonmetric sex determination from the distal and posterior humerus in black and white South Africans. *Journal of Forensic Sciences*, *56*, 710–714.

Vani, P., Malsawmzuali, J., Anbalagan, J., & Rajasekar, S. (2018). Morphological study of clavicular rhomboid impression in South Indian population with its clinical significance. *International Journal of Scientific Research*, *7*, 30–32.

Weiss, E. (2003). Understanding muscle markers: Aggregation and construct validity. *American Journal of Physical Anthropology*, *121*, 230–240.

Weiss, E. (2004). Understanding muscle markers: Lower limbs. *American Journal of Physical Anthropology*, *125*, 232–238.

Weiss, E. (2007). Muscle markers revisited: Activity pattern reconstruction with controls in a Central California Amerind population. *American Journal of Physical Anthropology*, *133*, 931–940.

Weiss, E., Corona, L., & Schultz, B. (2012). Sex differences in musculoskeletal stress markers: Problems with activity pattern reconstructions. *International Journal of Osteoarchaeology*, *22*, 70–80.

Wescott, D. J. (2000). Sex variation in the second cervical vertebra. *Journal of Forensic Sciences*, *45*, 462–466.

Wiredu, E. K., Kumoji, R., Seshadri, R., & Biritwum, R. B. (1999). Osteometric analysis of sexual dimorphism in the sternal end of the rib in a West African population. *Journal of Forensic Sciences*, *44*, 921–925.

Wrobel, G. D., Danforth, M. E., & Armstrong, C. (2002). Estimating sex of Maya skeletons by discriminant function analysis of long-bone measurements from the protohistoric Maya site of Tipu, Belize. *Ancient Mesoamerica*, *13*, 255–263.

Zumwalt, A. (2006). The effect of endurance exercise on the morphology of muscle attachment sites. *The Journal of Experimental Biology*, *209*, 444–454.

CHAPTER 9

Parturition markers and skeletal sex estimation

Clare McFadden
School of Archaeology and Anthropology, The Australian National University, Canberra, ACT, Australia

Introduction

The ability to identify whether an individual has been pregnant or has given birth has significant implications for our understanding of the individual and populations, and also has relevance to both forensic investigations and archaeological research. The estimation of parity status informs our understanding of sex, pregnancy, fertility, and childbirth. At an individual level and in a forensic context, it may assist with the identification of human skeletal remains. As a subset of sex in the biological profile, the identification of parturition markers may also be used as an indication of biological sex. In an archaeological context, it may add to the profile of an individual and provide the opportunity for us to better understand the female experience of pregnancy and childbirth in historical and prehistorical cultures. At a population level, it would give us an improved, empirical basis for estimating fertility rates and add to our understanding of infant and maternal mortality and maternal healthcare.

The human pelvis comprises four separate bones: two innominates, the sacrum, and the coccyx. The innominate has three primary elements, the ilium, ischium, and pubis, which fuse during early adolescence. The left and right innominates articulate with each other at the pubis, and with the sacrum at the auricular surface of the ilium. The sacrum is composed of sacral vertebrae, which also fuse during adolescence. In addition to the innominate, the sacrum articulates with the vertebral column and the coccyx. In subadults, the pelvis is largely of the same form in males and females. However, during pubertal growth, the female pelvis develops to accommodate childbirth. Coleman (1969) observed that much of the expansion of the pelvis occurs due to remodeling of the osseous elements of the pelvis, rather than metaphyses, through deposition and resorption, and described the process as dynamic, multidirectional, and complex (see Chapter 6 of this volume).

In addition to skeletal size and shape differences related to sex, a number of other features of the pelvis have been observed, seemingly in association with female sex and parturition status. These features are commonly termed parturition scars or markers, or pelvic scars, and are thought to result from a range of factors associated with pregnancy, including greater movement in the pelvis through relaxation of the ligaments, greater

https://doi.org/10.1016/B978-0-12-815767-1.00009-2

weight bearing, and trauma (e.g., strain, tearing) at ligament attachment sites through the birthing process. Two markers, dorsal pubic pitting (Figs. 1–3) and the preauricular sulcus or groove (Fig. 4), have traditionally defined parturition scars; however, studies have included additional features such as extension of the pubic tubercle (Bergfelder & Herrmann, 1980; Cox & Scott, 1992; Decrausaz, 2014; Maass & Friedling, 2016), height of the pubic tubercle (Decrausaz, 2014; Maass & Friedling, 2016; Snodgrass & Galloway, 2003), the interosseous groove of the ilium (Andersen, 1986; Houghton, 1974; Kelley, 1979; Maass & Friedling, 2016), and the iliac tuberosity (Andersen, 1986; Maass &

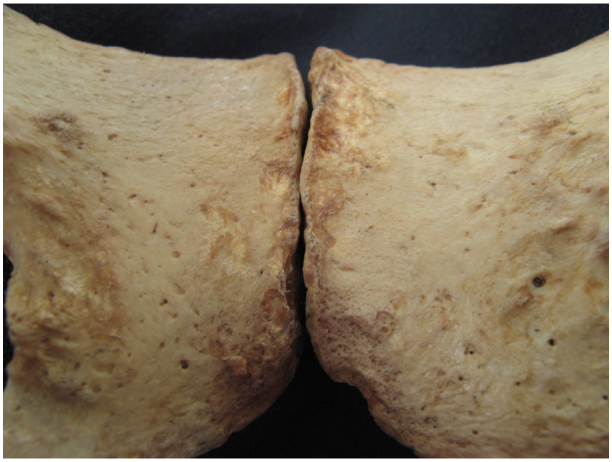


Fig. 1 Minimal or shallow dorsal pubic pitting. NMNH Terry 1616. *(Photos from the Terry Collection at the NMNH Physical Anthropology Division, taken by and courteously provided by Cheyenne Lewis.)*

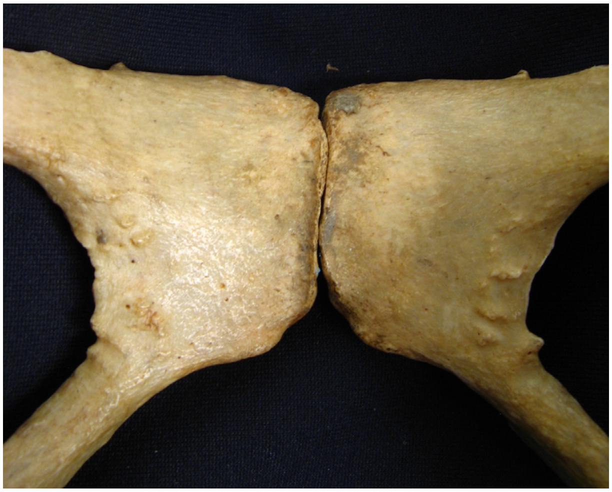


Fig. 2 Moderate dorsal pubic pitting. NMNH Terry 1608. *(Photos from the Terry Collection at the NMNH Physical Anthropology Division, taken by and courteously provided by Cheyenne Lewis.)*

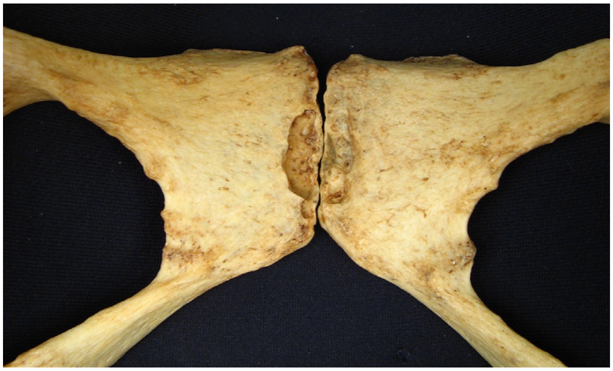


Fig. 3 Severe dorsal pubic pitting. NMNH Terry 1624. *(Photos from the Terry Collection at the NMNH Physical Anthropology Division, taken by and courteously provided by Cheyenne Lewis.)*

Fig. 4 A wide and deep preauricular groove, often referred to as the groove of pregnancy. *(Image courtesy of the Forensic Anthroplogy Center, University of Tennessee.)*

Friedling, 2016), and thus parturition scars may refer to any skeletal feature of the pelvis that has been found to be indicative of parity status. Notwithstanding, much of the research to date has focused on dorsal pubic pitting and the preauricular sulcus as the most promising features for reliable estimation of parturition and sex; and therefore, they are the focus of this chapter.

Origins of dorsal pubic pitting and the preauricular sulcus/groove

Walter G. Putschar first published his observations of the impacts of pregnancy and childbirth on the bony pelvis in his 1931 book *Development, Growth and Pathology of the Pelvic Joints of Man, with Special Emphasis on the Changes Caused by Pregnancy, Childbirth and Consequences*, which he later summarized in 1976. Putschar (1976) reported that remodeling on the dorsal aspect of the pubis, now known as dorsal pubic pitting, was frequently associated with positive parity status based on his personal observations. He attributed the cause of pitting to the loosening of ligaments that attach to the dorsal aspect of the pubis. Further modifications may have occurred due to hemorrhaging and tearing of these ligaments during childbirth and increased movement in the joints of the pelvis following childbirth (Putschar, 1976). He believed that the dimensions of the birth canal, size and cranial circumference of the infant, and age of the mother also impacted upon the manifestation of dorsal pubic pitting. Angel (1969) similarly observed the features described by Putschar (1931) in archaeological skeletal remains, and hypothesized that these features may increase in severity and frequency after multiple births.

Stewart (1970, p. 129) examined the presence of pelvic scarring in an archaeological Inuit sample and the Terry Collection of historic (birth years from 1828 to 1943) human skeletal remains. He reported the presence of remodeling in the form of "depressions, pits and/or cavities" on the dorsal aspect of the pubic bone in numerous females, ranging from trace to large. He stated that it was probable that pelvic scarring was caused by the trauma of childbirth due to the prevalence among females and absence in most males. He noted that there was some degree of variability in this relationship across and within samples, observing that scarring was less frequent in the Terry Collection and cited a number of examples where parous females had minimal or no scarring. Stewart (1970, p. 133) cautioned that pelvic scarring should be used with "extreme care" as an indication of parity until further studies were undertaken.

Houghton (1974) suggested that the preauricular groove or sulcus, located inferior to the auricular surface, could be utilized as an indicator of parity, with two distinct forms described as the groove of ligament and groove of pregnancy, being observed in males and nulliparous females, and parous females, respectively. The former was described as having an "even, flat floor," and the latter was said to be pitted (Houghton, 1974, p. 381). These definitions were based on observations from a skeletal collection without provenance or parity records; similar to Stewart (1970) and Houghton (1974) he seemed to have found the presence of the groove of pregnancy in some but not all females and absence in males—sufficient evidence for its association with parity. Further, he argued that the sacroiliac joint could be expected to bear greater strain than the pubic symphysis during pregnancy and childbirth, due to receiving more weight loading based on its position, and concluded that the preauricular groove offered a more reliable indicator of parity compared with dorsal pubic pitting.

Over the past several decades, a great number of evaluations and reevaluations of parturition markers have been undertaken, producing conflicting results regarding the relationship between various types of pelvic scarring, parity, and biological sex estimation. This is not altogether surprising when the high degree of variability among studies, including sample origins and methodologies, is considered.

Diverse sample origins and biological profile accuracy

Based on the composition of the samples, the research on parturition markers to date can be categorized into three study types: unknown sex and parity status; known parity status in a female sample; and known male and female samples with known parity status. Three studies utilized skeletal collections that lacked verifiable biological sex and parity data. Houghton (1974, p. 382) noted for his sample that the observer undertaking the examination of pelvic scarring also assigned sex to each pelvis based on "accepted morphological criteria." Much of Houghton's (1974) study is devoted to defining two distinct manifestations of the preauricular groove; and for this purpose, the lack of verifiable sex data is not an issue. However, Houghton's (1974) results and conclusions must be treated with caution: the classification of sex and that of the preauricular groove was not independent, and as such, the classified sex of individuals may have impacted upon the observer's evaluation of pelvic scarring features. Similarly, Ashworth, Allison, Gerszten, and Pezzia (1976) estimated the sex of individuals in their sample utilizing pelvic and cranial features (though the specific methodology is not reported) and burial goods (which may be related to gender rather than biological sex) prior to evaluating pelvic scarring, which potentially could have introduced bias (see Chapter 20 of this volume). Maass and Friedling (2016) estimated sex using the Buikstra and Ubelaker (1994) standards for the skeletal portion of their sample. As such, these studies suffer from an inbuilt error to the degree that is associated with the particular sex estimation technique applied, and this may undermine their analyses of the relationship between pelvic scarring, parturition, and sex.

A number of past studies have utilized female-only samples with known parity status, including Holt (1978), Kelley (1979), Suchey, Wiseley, Green, and Noguchi (1979), Bergfelder and Herrmann (1980), Cox and Scott (1992), and Snodgrass and Galloway (2003). Work by Kelley (1979), Bergfelder and Herrmann (1980), Cox and Scott (1992), and Snodgrass and Galloway (2003) focused on testing the hypothesis that pelvic scarring is associated with parturition, while Suchey et al. (1979) evaluated the relationship between scarring, parturition, and age. The major limitation of female-only studies is that if pelvic scarring were identified in males, it could be safely concluded that parturition is not the sole cause (if a cause at all). Thus, the exclusion of males from the sample restricts our understanding of the etiology, or etiologies, of pelvic scarring.

Andersen (1986), Arriaza and Merbs (1995), and Decrausaz (2014) used human skeletal collections that comprised both male and female individuals of known sex and parity status. By including males, these studies provide a far more comprehensive exploration of the etiology of pelvic scarring. Further to this, the use of radiographs for males and females, as per Dee (1981), Spring, Lovejoy, Bender, and Duerr (1989), and McArthur, Meyer, Jackson, Pitt, and Larrison (2016), provides the opportunity for observing change in a living sample; however, only Spring et al.'s (1989) sample included radiographs taken both pre- and postpregnancy. These were available for only a small subset of six females, but notably, there were no changes observed between the pre- and postpregnancy radiographs in any of these women.

The origins and size of the samples utilized in these studies is a major determinant of the accuracy of the associated sex, parity status, and medical history data. Samples of cadaveric or forensic origins (Bergfelder & Herrmann, 1980; Decrausaz, 2014; Holt, 1978; Kelley, 1979; Maass & Friedling, 2016; Snodgrass & Galloway, 2003; Suchey et al., 1979; Tague, 1990) and living samples (Dee, 1981; McArthur et al., 2016; Spring et al., 1989) have far more comprehensive associated records with regard to parity status, number of births, and medical history compared with archaeologically derived samples. Cox and Scott's (1992) and part of Decrausaz's (2014) samples were sourced from archaeological cemetery populations; Arriaza and Merbs (1995) and Ashworth et al. (1976) utilized mummified samples from Peru, while the origin of Houghton's (1974) sample is unclear. The quantity and quality of biological profile data for these studies is varied, but it must be acknowledged that there is greater potential for error in reconstructed and estimated details (including sex) based on traditional bioarchaeological methods than when using medical and legal records. Within the literature on parturition markers, samples of cadaveric and forensic origin tend to be of greater size than archaeological samples, and overall sample sizes in these studies have ranged from 49 to 486 individuals. As such, cadaveric and forensic samples offer more accurate biological profile data and sample sizes that are more viable for statistical analyses.

Methodological concerns

In the evaluation of dorsal pubic pitting, grading of severity has been the most common method employed, with scores ranging from two grades (cf. Ashworth et al., 1976; Tague, 1990) of expression up to five. Holt (1978), Kelley (1979), Suchey et al. (1979), Andersen (1986), Decrausaz (2014), and Maass and Friedling (2016) examined pelves for dorsal pubic pitting using a three-grade scale of "absent," "trace to small," or "slight" (e.g., Fig. 1) and "medium to large" or "moderate to severe" (e.g., Figs. 2 and 3); however, Suchey et al. (1979) pointed out the arbitrary nature of the categories. Arriaza and Merbs (1995) and McArthur et al. (2016) employed similar systems of four grades of severity. In an attempt to increase precision, Maass and Friedling (2016) categorized each grade by a quantitative range for the maximum diameter of the largest pit (trace to small <2.0 mm; medium to large >2.0 mm). Bergfelder and Herrmann (1980) and Snodgrass and Galloway (2003) utilized the Ullrich five-grade method. Bergfelder and Herrmann (1980) noted in their conclusions that the method was not found to be suitable with respect to the manifestation of pitting and suggested a three-stage system would be more practical. Suchey et al. (1979) made a similar observation; however, Snodgrass and Galloway (2003) did not appear to share this concern, noting no issues with the grading system. Cox and Scott (1992) assessed dorsal pubic pitting as absent or present; and where it was observed as present, they quantified the number of distinct pits. This would fail to take into account Ashworth et al.'s (1976) observation that after multiple births the series of pits and depressions may form a single large pit or groove. There appears to be disagreement among studies regarding the most appropriate way to categorize dorsal pubic pitting, what the differences in severity indicate, and whether they are valuable.

There is even more variation in the methodology for evaluating the preauricular groove, including descriptive categories, severity grades, and measurements of dimensions. This is potentially problematic when it comes to comparing studies. Houghton (1974) classified two forms of the preauricular groove: groove of parturition (GP) (e.g., Fig. 4) and groove of ligament (GL). He defined GP as being made up of pits that form a groove, smooth between ridges, and undulating; GL is described as being more variable—narrow (but sometimes wider), short, with an even flat floor. Kelley (1979), Andersen (1986), Cox and Scott (1992), and Decrausaz (2014) all utilized Houghton's (1974) categories but with variations: Kelley (1979) included GL in his three-grade model, along with his own categories of "broad-shallow" and "developed"; Cox and Scott (1992) and Decrausaz (2014) included two additional categories, and the former also included a four-grade severity measure and a quantitative measure of length and width; and Andersen (1986) included a five-grade severity measure. Dee (1981) examined pelves simply for the presence or absence of the preauricular groove, while Spring et al. (1989) graded the severity of the groove from one to five. Arriaza and Merbs (1995) and Maass and Friedling (2016) used a four-grade scale of severity; again, Maass and Friedling (2016) employed a quantitative measure first and a grade second, with each grade representing a range of width and length. Tague (1990) observed the presence or absence of bone resorption at the preauricular area.

There are other features of the pelvis that have been included sporadically across studies searching for indicators of parturition, with just as great variability in methodology. The pubic tubercle was measured by Snodgrass and Galloway (2003), Decrausaz (2014), and Maass and Friedling (2016), with the latter applying three grades to the measurement. Cox and Scott (1992) applied four grades of severity to the tubercle. Cox and Scott (1992) reported a correlation between the pubic tubercle and parity, but the remaining studies did not support this finding; and Decrausaz (2014) and Maass and Friedling (2016) found that a large tubercle was associated with male sex. Andersen (1986) applied five grades of severity and five types of shape to the iliac tuberosity, while Maass and Friedling (2016) applied three; neither reported significant results for the tuberosity and sex or parity. Snodgrass and Galloway (2003) and Decrausaz (2014) both measured tubercle distance and the angle of the arcuate line, with no significant findings relating to parity or sex. Kelley (1979) and Maass and Friedling (2016) applied the same grading systems they had used for the preauricular groove to the interosseous groove, with the latter study finding a weak association with sex, while Andersen (1986) used a seven-grade system and depth measurement and reported an association between sex, parity, and groove pitting. Arriaza and Merbs (1995) categorized a large range of features as absent, mild, moderate, or severe, including a raised facet on the iliac tuberosity, a protruding iliac spine, ventral pubic depression, sacroiliac ligament groove on the sacrum, depression on the sacral tuberosity, and facets on the sacral tuberosity, but none were found to be predictive of sex or parity. Cox and Scott (1992) observed the presence or absence of sulci along the anterior margins of the sacrum adjacent to the auricular facet, but found no association with parity.

There are a number of other methodological concerns plaguing research on parturition markers, primarily observer bias, asymmetry of features, arbitrary categorization of continuous features, and intra- and interobserver error. Several studies (cf. Cox & Scott, 1992; Snodgrass & Galloway, 2003) noted that observations were made blind to the sex records associated with the pelvis. However, for all studies, the location of pelvic scarring would make it difficult to undertake observations without also observing the pelvic features used to estimate sex, and this may have a significant impact upon the evaluation of features (Nakhaeizadeh, Dror, & Morgan, 2014). There is discrepancy between studies regarding the presence of asymmetry in the pelvic features of interest. Maass and Friedling (2016) noted no significant differences in the presence or absence of features in left and right os coxae and, on this basis, justified the use of the left only. This is in contrast to Houghton (1974), Kelley (1979), Bergfelder and Herrmann (1980), and Andersen (1986) who all observed some degree of variation by side. Houghton (1974) and Kelley (1979) elected to examine the side that exhibited the most severe scarring; Bergfelder and Herrmann (1980) noted their concern but did not state any change to methodology; and Andersen (1986) simply noted differences during data collection. Spring et al. (1989, p. 250) noted only "slight differences" in symmetry, but where it existed, they too graded the more severe side. It is clear that there are differing observations of the manifestation of pelvic scarring across the whole pelvis. This emphasizes the importance of understanding the histological changes that produce pelvic scarring.

A number of studies noted the arbitrary nature of grading systems applied to continuous features, and this is certainly a concern for methodologies throughout the field. Grades must be useful and reflective of the range of variation for each feature, as noted by Suchey et al. (1979) and Bergfelder and Herrmann (1980) in relation to the Ullrich method. Furthermore, grading systems can be difficult to replicate as they are highly subjective and are somewhat dependent on the range of variation within the sample being observed. Maass and Friedling (2016) attempted to overcome the issue of replicability by using a measurement that is then classified based on a set numerical demarcation point. However, one of the primary concerns with using quantitative methods is the loss of data in terms of overall appearance. In the case of dorsal pubic pitting, an overall assessment must be made of both the size and shape of individual pits and the quantity of pits. It would be very difficult to gain this information through a quantitative assessment. Observer error tests examine the replicability of a method, determining whether a method is of practical use to archaeologists and forensic anthropologists in the field. Intraobserver error refers to the rate of error in repeated attempts by a single observer; interobserver error refers to the rate of error between observers. Suchey et al. (1979) tested intraobserver error and found 14% of features were placed in different categories between the two attempts by a single observer. They reported that replicability increased when the bones were arranged in order of severity of the trait in question; Tague (1990) also employed this tactic prior to examining dorsal pubic pitting. Maass and Friedling (2016) tested inter- and intraobserver error using 30 randomly selected pelves. The differences were found to be nonsignificant with a maximum difference of 1.0 mm for the measurements of pelvic scarring. In contrast, Decrausaz (2014) reported a high rate of intraobserver error for measurements of the pelvis, but low rates for morphologically evaluated pelvic scarring. Snodgrass and Galloway (2003) noted no significant interobserver error; however, they did not report the exact rate or the means of testing. Spring et al. (1989) noted that separate observers examined each pelvis; and where there was disagreement between observers, they discussed and came to an agreed conclusion. While this was not a test of interobserver error, it is indicative that, at times, there were differences in opinion that may have contributed to interobserver error. McArthur et al. (2016) used the kappa agreement test to evaluate intraobserver error for dorsal pubic pitting and reported scores of 0.29–0.36 for presence, grading, and laterality, which is considered low agreement and indicative of a significant rate of error (McHugh, 2012).

Untangling correlation and causality

Most studies made some comment on the causal mechanism of pelvic scarring. Houghton (1974) stated that hormonal softening of ligaments during pregnancy was responsible for parturition scarring. He advised that osteoclastic resorption occurs adjacent to the ligamentous attachments as a result of increased strain and movement. Houghton (1974) suggested that the preauricular groove is more severely impacted than the dorsal aspect of the pubic bone due to the sacroiliac joint being in the direct line of transfer of body weight. Similarly, Kelley (1979, p. 541) noted that pelvic scarring occurs “at sites of ligamentous attachments” and suggested that the hormones released during pregnancy cause hypertrophy of the ligaments, which in turn causes resorption at attachment sites. He also argued that the subsequent tearing and hemorrhaging associated with childbirth could cause further changes. Kelley (1979) added that various factors, such as birth canal diameter, fetal head circumference, physical activity, number of pregnancies, obstetric care, and age at death, all impact the morphology of parturition scars. Bergfelder and Herrmann (1980) also noted the hormonal loosening of ligaments during pregnancy as potentially having some significance. Andersen (1986, p. 96) undertook an experimental approach to the pelvic flexibility hypothesis, describing a method whereby pelves were "articulated with rubber bands and a small 0.5 mm spacer between the pubic bones," following the methodology of Howells and Hotelling (1936). This was believed to represent the degree of flexibility in the skeletal pelvis. Snodgrass and Galloway (2003, p. 4) stated that hormonal levels during childbearing "may affect pelvic stability."

Holt (1978), Suchey et al. (1979), Bergfelder and Herrmann (1980), Andersen (1986), Spring et al. (1989), Cox and Scott (1992), Maass and Friedling (2016), and Decrausaz (2014) concluded that pelvic scarring is at least not solely caused by parturition. Spring et al. (1989) added that parturition may not cause pelvic scarring at all, based on observations from pre- and postpregnancy for six females. Similarly, Suchey et al. (1979) noted that eight females who were pregnant or very recently postpartum at the time of death did not display pelvic scarring. They suggested that "the basic mechanism of change needs to be studied by histological methods" (p. 523). The hypothesis that pelvic instability may cause scarring, which was argued to be a symptom of pregnancy by some studies, was also proposed by studies that did not consider scarring as a product of parity. Andersen (1986) suggested that pelvic scarring is caused by the flexibility of the pelvis, stating that females tend to have greater flexibility and, therefore, instability in the pelvis compared with males. She explained that the pelvis has a skeletal functional limit to flexibility and a muscular limit and that the skeletal and muscular systems work to stabilize the pelvis. She argued that if the skeletal functional limit and muscular limit are high, then excess motion may occur in the joints, subsequently causing resorption at ligament attachment sites. Andersen (1986) concluded that there is a causal relationship between pelvic flexibility and scarring. However, there is some doubt as to the validity of the test used to conclude that females have increased skeletal flexibility, as other studies have reported that males and nonpregnant females do not significantly differ in range of movement of the sacroiliac joint (Walker, 1992) and only differ slightly in mobility of the pubic symphysis (Walheim, Olerud, & Ribbe, 1984). Andersen's (1986) test of flexibility assumes that the flexibility of the skeletal pelvis reflects the flexibility of the muscular pelvis, which is inconsistent with similarities observed in the pelvic muscular flexibility of males and nonpregnant females. Maass and Friedling (2016) supported Andersen's (1986) hypothesis, stating that the female requirement for greater ligamentous stabilization may cause increased pelvic scarring. They concluded that the ratio of body size to pelvis size is a significant factor, arguing that smaller bodies with larger pelves tend to show more scarring. However, it is possible that this is another representation of the correlation with sex. Decrausaz (2014) evaluated age, occupation, height, weight, and pathology and concluded that correlations between pelvic scarring and body size and pelvis size in both sexes indicated that scarring has a musculoskeletal basis. Tague (1990) observed that humans and macaques both have large newborns compared to the pelvic inlet dimensions and concluded that the lack of resorption at the preauricular area of macaques suggests that resorption at this location in humans is not associated with parity. He suggested that resorption at the preauricular area may, alternatively, be due to bipedalism and difference in weight distribution. He concluded that there is a different cause for dorsal pubic pitting; estrogen is believed to induce resorption of the pubis in reproductively active females. Maass and Friedling (2016) also suggested that the causes of preauricular groove and dorsal pubic pitting are different, due to the lack of a significant relationship between the two.

It is possible that other biological factors such as age and pathological/health conditions may also obscure, interact with, or even cause pelvic scarring. Age data have been collected in numerous studies to understand its role in the manifestation of pelvic scarring. The pubic symphysis is known to modify with increasing age, and controlling for age is, therefore, particularly relevant with respect to the occurrence of dorsal pubic pitting. Kelley (1979), Suchey et al. (1979), Bergfelder and Herrmann (1980), Tague (1990), and Cox and Scott (1992) had associated age at death data for their samples and reported that increasing age had a significant impact on pelvic scarring, with some features increasing in severity with age (e.g., pubic tubercle extension as per Cox & Scott, 1992; and dorsal pubic pitting as per Suchey et al., 1979), and others seemingly remodeling and receding (Bergfelder & Herrmann, 1980; Kelley, 1979). Age was known for a part of Maass and Friedling's (2016) sample, but was estimated for the forensic component; they found that the pubic tubercle extension and iliac tuberosity increased in severity with age. Ashworth et al. (1976) and Arriaza and Merbs (1995) estimated the age of their samples using skeletal features, and also evaluated the health and status of the populations; the former did not analyze age, but the latter found the preauricular groove, iliac tuberosity, and sacral tuberosity to increase in severity with age. Snodgrass and Galloway (2003) evaluated age and body size/weight data and reported that weight became a significant predictor of scarring for females over 50 years of age, but no significant correlation between scarring and age was found. McArthur et al. (2016) obtained age data but only provided broad age categories in their paper. Bongiovanni (2016) reported that parous females often experience expedited aging of the pubic symphysis and, to a lesser extent, the auricular surface. Though complex, the literature to date indicates that age impacts upon the manifestation of at least some forms of pelvic scarring. The relationship between scarring and pathology has been less studied, perhaps as the prevalence of parturition scarring is not logically indicative of disease. Houghton (1974) reported pathology in four pelves, but these were removed from the study due to obstruction of the preauricular groove.

Holt (1978) noted that among the nulliparous women with scarring in his sample, three had inflammatory pelvic problems, one had a femoral hernia, seven had severe edema of one or both legs, and four were obese. Although not explicitly stated, Holt (1978) seems to imply that pathology and weight may be alternative causes of scarring. Ashworth et al. (1976) set out with the assumption that pelvic scarring is caused by parity, but found that there were significant differences between the two populations examined. They suggested that the differences between the two groups are likely due to different experiences of health: the colonial group lived in poverty, there was a large number of fractures suggesting ill-treatment and evidence of poor diet, while the preColumbian had abundant food offerings and lower manifestations of poor diet.

Metaanalysis: A quantitative approach to the current status of parturition scars

Based on the conflicting results of research to date, McFadden and Oxenham (2018) conducted a metaanalysis of parturition studies with sufficient data to reanalyze the relationship between dorsal pubic pitting and the preauricular groove, and parity and sex. Notably, studies that lacked known parity status and sex were excluded, due to the potential inaccuracy of reported results. Utilizing data from 11 studies, McFadden and Oxenham (2018) found a relationship between dorsal pubic pitting and parity, which notably lacked predictive strength, and a nonsignificant and negligible relationship between the preauricular groove and parity. Since that analysis was undertaken, only one study could be identified for further inclusion. McArthur et al. (2016) reported on the association of dorsal pubic pitting, identified on CT scan, with parity status. Their data, which they claimed to have demonstrated the association between dorsal pubic pitting and vaginal birth, have been included and the analyses of McFadden and Oxenham (2018) performed again.

Metaanalytic techniques can quantify, and to some extent clarify, the status of parturition markers, but also have the ability to highlight potential causes of conflicting results. In this iteration of the metaanalysis, the 11 studies identified in McFadden and Oxenham (2018) and the data from McArthur et al. (2016) were utilized and the characteristics of these are provided in Table 1, including the correct classification rates for each study. The two original hypotheses, whether parturition is responsible for scarring observed on the female pelvis and whether such scars are restricted to females, were tested. The results from all studies were normalized based on nonparous and parous status and absence and presence of indicators (dorsal pubic pitting and preauricular groove), with the absence of indicators expected to predict nonparity and male sex, and the presence on indicators expected to predict parous females. All grades above absent were considered indicators of parity. Individuals were considered correctly classified if they were assigned to the correct group (i.e., females with presence of scarring, males with absence of scarring, parous females with presence of scarring, and nonparous females with absence of scarring).

Table 1 Characteristics of studies included in the metaanalysis.

Study	Total (*n*)	Correctly classified	Proportion	95% confidence interval (exact)	
Dorsal pubic pitting and parity					
Holt (1978)	68	39	0.57	0.45	0.69
Kelley (1979)	198	122	0.62	0.54	0.68
Suchey et al. (1979)	486	355	0.73	0.69	0.77
Andersen (1986)	141	74	0.52	0.44	0.61
Cox and Scott (1992)	49	22	0.45	0.31	0.60
Decrausaz (2014)	125	65	0.52	0.43	0.61
McArthur et al. (2016)	311	236	0.76	0.71	0.81
Preauricular groove and parity					
Kelley (1979)	198	130	0.66	0.59	0.72
Andersen (1986)	141	73	0.52	0.43	0.60
Spring et al. (1989)	190	80	0.42	0.35	0.49
Cox and Scott (1992)	80	38	0.48	0.36	0.59
Decrausaz (2014)	138	64	0.46	0.38	0.55
Dorsal pubic pitting and sex					
Ashworth et al. (1976)	65	46	0.71	0.58	0.81
Andersen (1986)	226	159	0.70	0.64	0.76
Decrausaz (2014)	261	175	0.67	0.61	0.73
Maass and Friedling (2016)	304	217	0.71	0.66	0.76
McArthur et al. (2016)	359	249	0.69	0.64	0.74
Preauricular groove and sex					
Houghton (1974)	119	100	0.84	0.76	0.90
Dee (1981)	300	146	0.49	0.43	0.54
Andersen (1986)	226	193	0.85	0.80	0.90
Spring et al. (1989)	300	162	0.54	0.48	0.60
Decrausaz (2014)	286	185	0.65	0.59	0.70
Maass and Friedling (2016)	309	222	0.72	0.66	0.77

Modified from McFadden, C., & Oxenham, M. F. (2018). Sex, parity, and scars: A meta-analytic review. *Journal of Forensic Sciences, 63,* 201–206.

Metaanalysis is an effective tool that uses the combined power of multiple studies to calculate the correlation between variables (dorsal pubic pitting and the preauricular groove) and outcomes (parity and sex), or the size of the effect. It also allows us to examine the similarities and differences between studies and to identify studies that have produced significantly divergent results than others. As the additional study reported data solely for dorsal pubic pitting, only correlation and heterogeneity tests for pubic pitting and sex and parity were repeated. Statistical outputs included estimates of the overall success rate for relationships of interest, confidence intervals, and measures of heterogeneity (I^2). High levels of homogeneity among studies is desirable as it demonstrates consistency in results. In contrast, high heterogeneity may indicate significant differences in samples, methodologies, and analyses.

With the addition of the data from McArthur et al. (2016), the results for dorsal pubic pitting remained practically unchanged from those reported in McFadden and Oxenham (2018). The predictive capacity for parity increased from 64% to 66% (95% CI 64%–69%), while the relationship with sex remained unchanged at 70%. McFadden and Oxenham (2018) reported the predictive power of the preauricular groove to be 66% for sex and 52% for parity. The high level of heterogeneity among studies of the preauricular groove and sex and parity could be attributed to the lack of consistency in methodologies in this pool of studies. The homogeneity of the studies of dorsal pubic pitting and sex is reassuring that a genuine relationship exists between the two. It also calls into question whether the heterogeneity in dorsal pubic pitting and parity group can be attributed to differences in samples or methods, as we may expect that such factors would impact upon the heterogeneity of both groups equally.

Future research

If the cause of parturition markers is to be understood, the approach to their study needs to be revised. Samples of known sex, age, and parity status are essential to a robust study of pelvic scarring etiology. These variables, at a minimum, are related to each other both as a statement of fact (parturition status being dependent on sex) and as demonstrated in the literature (parturition status impacting upon aging of the pelvis and vice versa). As such, data for all three variables need to be obtained to fully explore the relationships. Additionally, the inclusion of both females and males, and subadults and adults, is necessary to our understanding of the breadth and diversity of the occurrence of scarring. Methodologies for evaluating scarring must be consistent, justified, and replicable, with low inter- and intraobserver error. Depending on the cause of scarring, the presence or absence may be a sufficient distinction. Of paramount importance, research questions need to be devised around cause rather than symptom. The literature on parturition scars contains a mix of pregnancy and childbirth-based hypotheses, and it is essential to distinguish between the two. Furthermore, the research to date has evaluated the correlation between scars and parity and sex, based on the assumption that scars are in some way (though not explicit in the research design) related to parity status. Correlations between these variables will only progress our understanding so far: if parity (either pregnancy or childbirth) or sex (mechanism not defined) are causal factors rather than associations, then the causal hypothesis needs to be tested.

Conclusion

In their article "Skeletal Indicators of Pregnancy and Parturition: A Historical Review," Ubelaker and De La Paz (2012, p. 870) concluded that "the published literature presents contradictory evidence and discussion relating to the issue of skeletal alterations and their association with childbirth." Based on research to date, pelvic scarring cannot be considered a reliable indicator of parity, and there are alternative, more accurate methods of estimating sex. Nonetheless, there is clearly an association between parturition markers and parity and sex, even if it is not a strongly predictive one. The question then remains as to what causes these scars, and it is suggested that a new approach to the study of parturition markers is required. The biological mechanisms behind parturition markers require greater consideration and hypothesis testing in order to progress our understanding of their etiology.

References

Andersen, B. (1986). *Parturition scarring as a consequence of flexible pelvic architecture* [dissertation]. Vancouver, BC: Simon Fraser University

Angel, J. L. (1969). The bases of paleodemography. *American Journal of Physical Anthropology*, *30*(3), 427–437.

Arriaza, B., & Merbs, C. F. (1995). Evidence of childbirth in the pelvis of prehistoric Andean women. *Estudios de Antropología Biológica*, *5*(1), 65–79.

Ashworth, J., Allison, M., Gerszten, E., & Pezzia, A. (1976). The pubic scars of gestation and parturition in a group of pre-Columbian and Colonial Peruvian Mummies. *American Journal of Physical Anthropology*, *45*(1), 85–90.

Bergfelder, T., & Herrmann, B. (1980). Estimating fertility on the basis of birth-traumatic changes in the pubic bone. *Journal of Human Evolution*, *9*, 611–613.

Bongiovanni, R. (2016). Effects of parturition on pelvic age indicators. *Journal of Forensic Sciences*, *61*, 1034–1040.

Buikstra, J. E., & Ubelaker, D. (1994). *Standards for data collection from human skeletal remains*. Research Series No. 44. Fayetteville, Arkansas.

Coleman, W. H. (1969). Sex differences in the growth of the human bony pelvis. *American Journal of Physical Anthropology*, *31*(2), 125–151.

Cox, M., & Scott, A. (1992). Evaluation of the obstetric significance of some pelvic characters in an 18th century British sample of known parity status. *American Journal of Physical Anthropology*, *89*, 431–440.

Decrausaz, S. (2014). *A morphometric analysis of parturition scarring on the human pelvic bone* [dissertation]. Victoria, BC: University of Victoria.

Dee, P. M. (1981). The preauricular sulcus. *Radiology*, *140*(2), 354.

Holt, A. (1978). A re-examination of parturition scars on the human female pelvis. *American Journal of Physical Anthropology*, *49*, 91–94.

Houghton, P. (1974). The relationship of the pre-auricular groove of the ilium to pregnancy. *American Journal of Physical Anthropology*, *41*, 381–390.

Howells, W., & Hotelling, H. (1936). Measurements and correlations on pelves of Indians of the Southwest. *American Journal of Physical Anthropology*, *21*(1), 91–106.

Kelley, M. (1979). Parturition and pelvic changes. *American Journal of Physical Anthropology*, *51*, 541–546.

Maass, P., & Friedling, L. (2016). Scars of parturition? Influences beyond parity. *International Journal of Osteoarchaeology* *26*(1), 121–131.

McArthur, T. A., Meyer, I., Jackson, B., Pitt, M. J., & Larrison, M. C. (2016). Parturition pit: The bony imprint of vaginal birth. *Skeletal Radiology*, *45*(9), 1263–1267.

McFadden, C., & Oxenham, M. F. (2018). Sex, parity, and scars: A meta-analytic review. *Journal of Forensic Sciences*, *63*, 201–206.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

Nakhaeizadeh, S., Dror, I. E., & Morgan, R. M. (2014). Cognitive bias in forensic anthropology: Visual assessment of skeletal remains is susceptible to confirmation bias. *Science & Justice*, *54*(3), 208–214.

Putschar, W. G. (1931). *Development, growth and pathology of the pelvic joints of man, with special emphasis on the changes caused by pregnancy, childbirth and consequences*. Jena: Verlag von Gustav Fischer.

Putschar, W. G. (1976). The structure of the human symphysis pubis with special consideration of parturition and its sequelae. *American Journal of Physical Anthropology*, *45*, 589–594.

Snodgrass, J., & Galloway, A. (2003). Utility of dorsal pits and pubic tubercle height in parity assessment. *Journal of Forensic Science*, *48*(6), 1226–1230.

Spring, D., Lovejoy, C. O., Bender, G., & Duerr, M. (1989). The radiographic preauricular groove: Its non-relationship to past parity. *American Journal of Physical Anthropology*, *79*, 247–252.

Stewart, T. (1970). Identification of the scars of parturition in the skeletal remains of females. In T. Stewart (Ed.), *Personal identification in mass disasters* (pp. 127–134). Washington (DC): National Museum of Natural History.

Suchey, J., Wiseley, D., Green, R., & Noguchi, T. (1979). Analysis of dorsal pitting in the Os pubis in an extensive sample of modern American females. *American Journal of Physical Anthropology*, *51*, 517–540.

Tague, R. (1990). Morphology of the pubis and preauricular area in relation to parity and age at death in *Macaca mulatta*. *American Journal of Physical Anthropology*, *82*(4), 517–525.

Ubelaker, D., & De La Paz, J. (2012). Skeletal indicators of pregnancy and parturition: A historical review. *Journal of Forensic Sciences*, *57*(4), 866–872.

Walheim, G., Olerud, S., & Ribbe, T. (1984). Mobility of the pubic symphysis: Measurements by an electromechanical method. *Acta Orthopaedica Scandinavica*, *55*, 203–208.

Walker, J. (1992). The sacroiliac joint: A critical review. *Physical Therapy*, *72*(12), 903–915.

CHAPTER 10

Dentition in the estimation of sex

Marin A. Pilloud, G. Richard Scott
Department of Anthropology, University of Nevada, Reno, NV, United States

There are multiple ways to assess differences between males and females in the dentition. For example, differences have been identified in the length of the tooth root (Harris & Couch, 2006; Zorba, Vanna, & Moraitis, 2014), diagonal measurements of the crown (Manchanda, Narang, Kahlon, & Singh, 2015; Peckmann, Meek, Dilkie, & Mussett, 2015), and intercusp distances (Harris & Dinh, 2006). Ratios of the canine have been used to identify sex differences (Bakkannavar, Manjunath, Nayak, & Pradeep Kumar, 2015; Iqbal, Zhang, & Mi, 2015), but others have found limited utility in the mandibular canine index (Acharya, Angadi, Prabhu, & Nagnur, 2011; Acharya & Mainali, 2009; Silva et al., 2016). The absolute volume of the canine has also been used to identify sex (De Angelis et al., 2015) as has the angle of the canine cusp (Calhoun, Guatelli-Steinberg, & Hubbe, 2018). Ratios of dentine and enamel (Garcia-Campos et al., 2018) and overall tooth weight (Schwartz & Dean, 2005) are also sexually dimorphic.

Within this treatment, the focus is primarily on sexual dimorphism of maximum crown dimensions with some discussion of dental morphology. These aspects of the dentition were chosen as there is much published research on these topics, and they are most readily relevant to the field of forensic anthropology. Before these topics are discussed, a review of dental development and evolution is presented in the context of sexual dimorphism. While much work has been done studying sexual dimorphism of deciduous teeth (e.g., Adler & Donlon, 2010; Black, 1978; Cardoso, 2010; Chan, 2007; DeVito & Saunders, 1990; Harris & Lease, 2005; Kondoh, 1990; Viciano, López-Lázaro, & Alemán, 2013), this chapter focuses primarily on the permanent dentition.

Dental development

Teeth comprise three main tissues: enamel, dentin, and cementum. Teeth begin to develop in utero within the developing maxilla and mandible. Development is initiated at the dentine horn with the deposition of dentine. Enamel is then laid down over the developing dentine. This deposition progresses from the occlusal tips of the tooth to the cement-enamel junction, where cementum then covers the dentine (instead of enamel) down the root of the tooth to the apex. This process of odontogenesis is recorded in the

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00010-9

microstructures of the enamel and dentin. Enamel is laid down in daily increments that are visible as cross-striations in histological sections of the tooth, making it possible to reconstruct the timing and duration of growth (Nanci, 2003).

Humans are dyphodont, meaning they have two sets of teeth (deciduous and permanent). Both sets of dentition are formed in the same manner and erupt through the bone and gingiva as teeth develop. Among permanent teeth, the first molar begins the sequence of eruption. This pattern holds for all primates (Smith, 1994). Following the first molar, eruption sequences are mesial to distal beginning with the incisors. The canine erupts around the same time as the premolars, and the second and third molars are the last teeth to erupt (Scheid & Weiss, 2012). Except for the third molar, female dental development is ahead of males (Heim, 2018), which is also generally true of skeletal growth.

The majority of tooth crowns and roots develop prior to puberty, which begs the question of the role of hormones in dental sexual dimorphism. There is some evidence to suggest that intrauterine hormones affect tooth size. In particular, research has found that females in dizygotic opposite-sex twins have larger crown dimensions on average, which may be related to the intrauterine testosterone of their twin brother (Ribeiro, Brook, Hughes, Sampson, & Townsend, 2013). Conversely, some researchers have argued that sexual dimorphism is more pronounced in later developing teeth in response to developing hormones during puberty (Gingerich, 1974; Kondo & Townsend, 2004; Kondo, Townsend, & Yamada, 2005). However, work by Guatelli-Steinberg, Sciulli, and Betsinger (2008) looked at sexual dimorphism in the mesiodistal measurements of several populations to test if later forming teeth were more subject to hormonal influences directing sexual dimorphism. Their results did not find later forming teeth to be more dimorphic, leading them to conclude that hormones are not the main contributor to sexual dimorphism of teeth. There are likely multiple causal factors, which could include the sex chromosomes and genes controlling the length of dental development, among many other biological reasons.

The structural gene for amelogenin, the proteins involved in amelogensis, is located on the X and Y chromosomes (Lau, Mohandas, Shapiro, Slavkin, & Snead, 1989). Alvesalo (1997, 2009, 2013) found that the Y chromosome has a role in enamel and dentine development, whereas the X chromosome is largely confined to enamel production. This pattern is observed in the differences between male and female teeth, as males consistently have more dentine than females, resulting in larger teeth (Garcia-Campos et al., 2018; Saunders, Chan, Kahlon, Kluge, & FitzGerald, 2007; Schwartz & Dean, 2005; Smith, Olejniczak, Reid, Ferrell, & Hublin, 2006). This pattern is also evident in the deciduous dentition (Harris, Hicks, & Barcroft, 2001).

As chromosomes are involved in odontogenesis, chromosomal abnormalities involving the sex chromosomes influence crown and root morphology. Individuals with Klinefelter syndrome (47, XXY) have larger crown sizes due to thick layers of enamel

(Alvesalo, Tammisalo, & Townsend, 1991) as well as longer roots (Lähdesmäki & Alvesalo, 2007) and larger molar cusp volume and height (Mayhall, Alvesalo, & Townsend, 1998). Some research indicates a higher rate of taurodontism, expansion of the pulp chamber (Schulman et al., 2005; Varrela & Alvesalo, 1988), and malocclusion (Alvesalo & Laine, 1992; Laine, Alvesalo, & Lammi, 1992) among these individuals. Similarly, individuals with Turner's syndrome (XO or 45,X) have smaller crown and root sizes (Lähdesmäki & Alvesalo, 2006). Kirveskari and Alvesalo (1982) found XO individuals have reduced frequencies and expressions of shoveling, especially on the maxillary second incisor, the hypocone of the maxillary first and second molars, the hypoconulid of the mandibular first and second molars, and Carabelli's trait. In a Finnish sample, Tomes' root and two-rooted lower premolars were more common in individuals lacking an X chromosome (23%–25%) than in relatives (2%–3%) (Varrela, 1992).

Primates and evolution

The canine is the most sexually dimorphic tooth among primates. Humans are on the low end of the canine sexual dimorphism scale, with male canines being around 6% larger than female canines (Garn, Kerewsky, & Swindler, 1966; Garn, Lewis, & Kerewsky, 1964). At the other extreme is *Papio anubis*, the Anubis baboon, with a maxillary male canine that exceeds that of females by up to nearly 80% (Garn, Kerewsky, & Swindler, 1966). In many primates, the canine is the last tooth to erupt. Among great apes, the canine emerges last along with the third molar (Swindler, 2002). Larger tooth sizes may be related to prolonged development; in fact, permanent canine crowns take longer to form in male apes than females apes (Schwartz & Dean, 2001), and ape canines overall take longer to develop than they do in humans (Swindler, 2002).

Among primates, body and canine size dimorphism are linked to agonistic behaviors (Plavcan, 2001). Plavcan (2012) argues that selection for larger female size over time leads to a reduction in sexual dimorphism over the course of human evolution, and that this selection could be related to fecundity, available resources, and provisioning of offspring. The reduction in canine dimorphism occurs early in hominin evolution, as some have argued that *Australopithecus afarensis* had canine dimorphism similar to that of modern humans (Kimbel & Delezene, 2009; Leutenegger & Shell, 1987).

Brace and Ryan (1980) argued that continued dental dimorphism among modern humans was related to body dimorphism retained into the Middle Pleistocene. In the Late Pleistocene, changes in food processing led to dental dimension reductions, but sexual dimorphism was retained until the end of the Pleistocene when hunting technology changed. Therefore, those populations that are *most* distant from large game hunting should be the *least* dimorphic. There is also an overall associated reduction in dental size after the adoption of agriculture (Pinhasi & Meiklejohn, 2011).

Sexual dimorphism of dental morphology

In dental morphology, trait definitions and scores vary across studies. However, studies typically follow the definitions of Turner, Nichol, and Scott (1991) and Scott and Irish (2017). The reader is referred to these publications for trait definitions. For decades, researchers have tested for sex differences in tooth morphology. Reports of significant sex differences for specific traits vary. For Carabelli's trait, some researchers found significant sex differences (e.g., Goose & Lee, 1971; Joshi, Godiawala, & Dutia, 1972; Kaul & Prakash, 1981; Kieser & Preston, 1981; Mizoguchi, 1985; Scott, Potter, Noss, Dahlberg, & Dahlberg, 1983; Townsend & Brown, 1981), but others reported no difference between males and females (e.g., Bang & Hasund, 1972; Garn, Kerewsky, & Lewis, 1966; Scott, 1980; Townsend, 1992; Turner, 1969). Mizoguchi (1985) presents frequency distributions for 12 crown traits in two Japanese samples. In tests for sex dimorphism, 6 of 24 comparisons show significant male-female differences. In four of six cases, one sample shows a significant sex difference, while the other does not. Only Carabelli's trait showed a significant sex difference in both samples. This suggests that sampling error plays a significant role in reports of sex dimorphism in morphology by disparate researchers studying diverse samples.

When significant male-female differences are reported for crown traits, males typically show higher frequencies and more pronounced expressions. For upper central incisors, Harris (1980) reports that females exhibit more shoveling than males. In six geographic samples, females show higher frequencies of shoveling in five instances. Except for Polynesians and Melanesians, the differences in shoveling between males and females are non-significant. When lingual fossa depth is used as a measure of shoveling, researchers found either a significant sex difference in favor of females (Rothhammer, Lasserre, Blanco, Covarrubias, & Dixon, 1968) or no sex difference (Aas, 1979; Mizoguchi, 1985). Kirveskari and Alvesalo (1981) reported greater lingual fossa depth in 13 Finnish 47, XYY males than in male and female relatives for upper first and second incisor measurements, but the differences were only significant for upper second incisor shoveling. Overall, the case for sexual dimorphism in shoveling is not well established. The allele EDAR V370A has a significant effect on shoveling expression and frequencies (Kimura et al., 2009; Park et al., 2012) and it is autosomal, not X-linked. Even if there is eventual agreement on the nature of shoveling sexual dimorphism, the difference between males and females is subtle and would be of limited utility in a forensic context.

The crown trait that shows consistent sexual dimorphism in diverse samples is the distal accessory ridge of the upper and lower canines (Fig. 1) (Abrantes, Santos, Pestana, & Pereira, 2015; Kaul & Prakash, 1981; Kieser & Preston, 1981; Scott, 1977; Scott et al., 1983; Scott & Turner, 1997). Sex dimorphism in the mesiodistal diameter of the lower canine is among the highest of all human tooth dimensions (Garn, Cole,

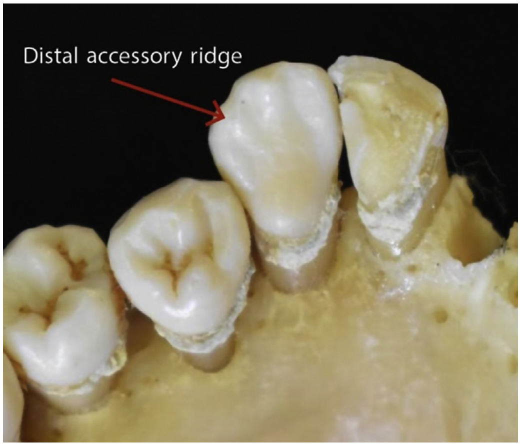


Fig. 1 Distal accessory ridge of the lower left canine.

Wainwright, & Guire, 1977); however, see below for more discussion. There is a modest but significant correlation between the expression of the distal accessory ridge and the mesiodistal diameter of the lower canine. When Noss, Scott, Potter, Dahlberg, and Dahlberg (1983) estimated the contribution of tooth size to sex dimorphism in distal accessory ridge expression on the lower canine, they found a male-female dimorphism of 75%. Controlling for crown size, the dimorphism was reduced to 47%. Tooth size plays a role in this sex difference, but other factors are at work.

In a large Pima Native American sample (approximately 1200 males and females), there was a significant sex difference in eight traits exhibited on nine teeth (Scott et al., 1983). Originally, the chi-square statistic was used to evaluate male-female differences. As an alternative to this method, ANOVA determined if mean trait scores were significantly different between males and females (Table 1). Incisor winging showed a significant sex difference, but was not scored on a ranked scale, so this variable is not included in the table. All other variables were scored on scales that had five to eight grades of expression. Shoveling of the upper central incisor showed a significant sex difference when evaluated by chi-square, but mean trait expression (MTE) is almost identical for males and females. Cusp 7 differed between males and females based on chi-square, but MTE values are similar and non-significant. The remaining six traits are significantly different through both lines of analysis. As expected, the lower canine distal accessory ridge is the most distinctly dimorphic trait whether analyzed by scales or male-female differences in MTE values (0.62). The hypocone of UM2 is the second most dimorphic trait (0.42). The other traits that show significant differences differ in MTE by relatively modest values (0.08–0.25).

Even though genes on the sex chromosomes are involved in dental development, crown and root traits show little or no sexual dimorphism at the phenotypic level.

Table 1 Crown traits that exhibit sexual dimorphism in a large Pima Native American sample.

Trait	Tooth	df	χ^2	*P*	MTE Male	MTE Female	Male-female difference	ANOVA *P*
Shoveling	UI1	5	11.83	.04	3.24 [595]	3.25 [659]	−0.01	N.S.
Distal accessory ridge	LC	4	71.34	<.001	1.55 [379]	0.93 [424]	0.62	<.05
Carabelli's trait	UM1	4	23.48	<.001	1.31 [609]	1.09 [633]	0.22	<.05
Hypocone	UM1	2	9.09	.01	3.94 [608]	3.86 [656]	0.08	<.05
Hypocone	UM2	4	19.65	<.001	2.61 [344]	2.19 [407]	0.42	<.05
Protostylid	LM1	4	9.56	.05	0.81 [532]	0.63 [554]	0.18	<.05
Cusp 6	LM1	3	15.28	.00	1.06 [532]	0.81 [612]	0.25	<.05
Cusp 7	LM1	4	10.36	.04	0.49 [619]	0.47 [628]	0.02	N.S.

MTE, mean trait expression. By convention, U = upper, L = lower, I = incisor, C = canine, M = molar, and the number represents the tooth position.

When differences are found, they are usually inconsistent among samples and low order in magnitude. For this reason, crown and root traits share with autosomal genetic traits the advantage that male and female data can be pooled to estimate population frequencies. This advantage is critical in the analysis of small skeletal samples where subdivision by sex often results in intolerably small samples. As numerous papers have demonstrated, there is no sex difference in all but one trait (canine distal accessory ridge). It is now common for workers to combine male and female data, no longer testing for sex differences (e.g., Coppa, Cucina, Mancinelli, Vargiu, & Calcagno, 1998; Cucina, Lucci, Vargiu, & Coppa, 1999; Irish, 2006; Manabe et al., 2003; Matsumura, 2007; Scott, Anta, Schomberg, & De La Rue, 2013).

Sexual dimorphism of dental size

In measurements of the skeleton, male and female humans can show a size difference nearing 20% (Humphrey, 1998). However, the teeth show a much smaller level of sexual dimorphism, at around 3%–7% (Harris & Foster, 2015), depending on the tooth, measurement, and population under study. Crown measurements are typically taken of the enamel in two planes: mesiodistal and buccolingual (also termed faciolingual or labiolingual). In the buccolingual plane, the dimension is the maximum width of the crown. For mesiodistal dimensions, measurements are taken as the maximum length or as the distance between the interproximal facets. The maximum dimensions follow the definition of Moorrees and Reed (1964), while those based on contact facets follow the definitions of Moorrees (1957) and Pedersen (1949). See also the discussion in Mayhall (1992) regarding these measurements. Finally, measurements can be taken of the cervix of the crown. These measurements are gaining popularity as they increase sample sizes when teeth are worn, broken, or have large carious lesions. Another advantage is that these dimensions are not impacted by the presence or absence of morphological crown traits. The definitions of these measurements follow Hillson, FitzGerald, and Flinn (2005) for permanent teeth, and Pilloud and Hillson (2012) for deciduous teeth.

While a few studies have documented low levels of dimorphism in tooth size (Kaur & Chattopadhyay, 2003; Angadi, Hemani, Prabhu, & Acharya, 2013; Peckmann, Logar, Garrido-Varas, Meek, & Pinto, 2016), significant dimorphism has been found in samples from South Africa (Macaluso, 2011), Turkey (İşcan & Kedici, 2003), Spain (Viciano et al., 2013), Japan (Adams & Pilloud, 2019), Brazil (Martins Filho, Lopez-Capp, Biazevic, & Michel-Crosato, 2016), Greece (Mitsea, Moraitis, Leon, Nicopoulou-Karayianni, & Spiliopoulou, 2014), India (Anuthama et al., 2011; Jain, Garg, Singh, Ansari, & Sangamesh, 2011; Schwartz & Dean, 2005), Libya (El Sheikhi & Bugaighis, 2016), Jordan (Shaweesh, 2017), Nepal (Acharya & Mainali, 2007), Egypt (Saikiran et al., 2014), Malaysia (Khamis, Taylor, Malik, & Townsend, 2014), and Portugal (Pereira, Bernardo, Pestana, Santos, & de Mendonça, 2010). These studies generally focus on tooth crowns, but some have explored sexual dimorphism of cervical measurements (Adams & Pilloud, 2019; Hassett, 2012; Kazzazi & Kranioti, 2018; Tuttösí & Cardoso, 2015; Zorba, Moraitis, & Manolis, 2011).

Studies on dental sexual dimorphism have found different teeth and measurements to show varying levels of dimorphism, but the canine is typically the most dimorphic. Garn, Kerewsky, and Swindler (1966) argued for a canine field effect in sexual dimorphism, stating that teeth on either side of the canine would also show greater dimorphism than the tooth in the same field further from the canine. For example, the lateral incisor and the third premolar would be more dimorphic than their counterparts further from the canine (i.e., the central incisor and the fourth premolar and molars). This theory has not been thoroughly tested, nor does it offer a means to account for the dimorphism of molars and the low dimorphism typically assigned to incisor teeth. For example, Pilloud, Hefner, Hanihara, and Hayashi (2014) found in a large global sample that all crown measurements exhibited sexual dimorphism *except* for the upper second incisor and both lower incisors.

To explore the various levels of dimorphism in different teeth and measurements, the same dataset used in Pilloud et al. (2014) was revisited. This dataset represents around 5600 individuals from around the globe dating to approximately the last 500 years (Table 2). All data were collected by T. Hanihara and are outlined in his publications (e.g., Hanihara, 1998; Hanihara & Ishida, 2005). Dental measurements were subject to an ANOVA test and an independent sample *t*-test (not assuming equal variance). Sexual dimorphism was calculated according to Garn, Lewis, Swindler, and Kerewsky (1967), where the male-to-female ratio is expressed as a percentage (male/female − 1.00). Positive values indicate males are larger; negative values indicate females are larger. These analyses were performed in SPSS v. 24 (SPSS Inc, 2016).

Table 2 Summary of dental metric data used in statistical analyses, based on Pilloud et al. (2014).

Broad geographical grouping	Regional group	Location	F	M	TOTAL
AFRICA (*n* = 858)	East Africa	Kenya, Somalia, Tanzania, Uganda	42	304	346
	Sub-Saharan Africa	Cameroon, Congo, Ethiopia, Gabon, Gambia, Ghana Ashanti, Guinea, Ivory Coast, Lesotho, Malawi, Mozambique, Ruwanda, South Africa Bushman, South African Hottentot, South Africa Kaffir, South Africa Zulu, Zambia, Zimbabwe	33	282	315
	West Africa	Liberia, Niberia, Senegal, Sierra Leone	17	165	182
	Guyana	Guyana	0	8	8
	Jamaica	Jamaica	0	7	7
ASIA (*n* = 3718)	Melanesia	Bismark, Fiji, New Britain, New Caldonia, New Hebrides, New Ireland, Papua New Guinea, Santa Cruz, Solomon, Torres Strait	277	618	895
	Micronesia	Caroline Islands, Caroline Ponape, Caroline, Gilbert Islands, Mariana Saipan, Mariana Tinian, Marshall Islands	25	73	98
	Native American	Alabama, Alaska, Arch Lake, Arizona, Arkansas, California, Colorado, Delaware, Florida, Georgia, Horn Shelter, Illinois, Kansas, Kentucky Indian Knoll, Kentucky, Louisiana, Maryland, Michigan, Mississippi, Missouri, Montana, Nebraska, Nevada, New Jersey, New Mexico, New York, North Dakota, Ohio, Oregon, Pennsylvania, South Dakota, Tennessee, Texas, Utah, Virginia, Washington, West Virginia, Wisconsin, Wyoming	294	433	727
	Polynesia	Chatham Islands Moriori, Cook Islands, Easter Islands, Gambier Islands, Hawaii, Marquesas, New Zealand Maori, Samoa, Society Islands, Tonga, Tuamotu Islands	252	718	970

Table 2 Summary of dental metric data used in statistical analyses, based on —cont'd

Broad geographical grouping	**Regional group**	**Location**	**F**	**M**	**TOTAL**
	South East Asian	Bali, Borneo, Cambodia, Celebes, Java, Laos, Lesser Sunda, Maccassar, Malacca, Malay, Molucca, Myanmar, Negrito Phillipines, Negrito Semang, Nicobar Islands, Phillipines, Sulu, Sumatra, Sumbawa, Thailand, Timor, Vietnam	86	694	780
	East Asian	China, Japan, Korea	51	197	248
EUROPE ($n=1055$)	Europe	Albania, Austria, Belgium, Bulgaria, Czecho, Denmark, Finland, France, Germany, Greece, Herzegovina, Holland, Hungary, Italy, Lapp, Norway, Poland, Portugal, Romania, Russia, Spain, Sweden, Switzerland, Yugoslavia	123	635	758
	Spitalfields	Spitalfields	102	195	297
Total			1302	4329	5631

With no consideration for population, the summary statistics, dimorphism, and results of ANOVA and *t*-tests for all males and all females are presented in Table 3. Again, all measurements are significantly different between males and females except for both lower incisors and the upper lateral incisor. The most dimorphic measurements are the buccolingual dimensions of the lower and upper canine. These are followed by (in order of most to least dimorphism) buccolingual upper third molar, mesiodistal lower canine, buccolingual upper second molar, mesiodistal lower third molar, buccolingual lower third premolar, buccolingual lower third molar, and buccolingual upper fourth premolar.

While general trends in tooth size dimorphism are clear in this large sample, when placing an individual in forensic casework, it is also important to consider population variation. Tooth size has been found to vary significantly between populations (e.g., Brook et al., 2009; Hanihara & Ishida, 2005; Pilloud et al., 2014; Schnutenhaus & Rösing, 1998). Using this same large dataset, data were divided by population and sex for all measures. The data were subject to linear discriminant function analysis, and the canonical variates were plotted in a scattergram to illustrate sample distribution in relation to sex and population (Fig. 2). This analysis was completed in the R computing environment (R Core Team, 2013). The scattergram shows some separation by group and sex;

Table 3 Summary statistics for all individuals divided by sex.

	Male				Female				% Dimorphism		
	N	Mean	St. dev	Std. error mean	*N*	Mean	St. dev	Std. error mean	(M/F − 1.00)	ANOVA	*t*-test
UI1_MD	992	8.69	0.58	0.0186	378	8.55	0.56	0.0289	1.70	0.000	0.000
UI2_MD	1199	7.14	0.66	0.0189	474	7.08	0.62	0.0286	0.85	0.086	0.079
UC_MD	1867	8.00	0.52	0.0121	648	7.81	0.50	0.0196	2.40	0.000	0.000
UP3_MD	2512	7.32	0.52	0.0105	759	7.17	0.51	0.0183	2.07	0.000	0.000
UP4_MD	2508	6.97	0.53	0.0105	754	6.86	0.50	0.0183	1.56	0.000	0.000
UM1_MD	3292	10.89	0.63	0.0110	1051	10.68	0.65	0.0201	1.94	0.000	0.000
UM2_MD	3153	10.25	0.74	0.0131	987	10.00	0.72	0.0229	2.52	0.000	0.000
UM3_MD	2040	9.34	0.84	0.0185	604	9.24	0.80	0.0325	1.04	0.012	0.010
LI1_MD	908	5.44	0.36	0.0119	370	5.44	0.36	0.0189	−0.14	0.730	0.732
LI2_MD	1177	6.09	0.43	0.0127	466	6.10	0.45	0.0208	−0.09	0.814	0.817
LC_MD	1508	7.09	0.51	0.0132	523	6.83	0.48	0.0208	3.81	0.000	0.000
LP3_MD	1898	7.20	0.54	0.0124	621	7.01	0.50	0.0202	2.64	0.000	0.000
LP4_MD	1875	7.30	0.57	0.0131	624	7.18	0.54	0.0214	1.68	0.000	0.000
LM1_MD	2281	11.62	0.65	0.0137	722	11.41	0.66	0.0246	1.85	0.000	0.000
LM2_MD	2339	11.13	0.80	0.0165	750	10.86	0.77	0.0281	2.49	0.000	0.000
LM3_MD	1863	11.13	0.94	0.0219	533	10.74	0.96	0.0416	3.67	0.000	0.000
UI1_BL	1123	7.38	0.50	0.0149	404	7.23	0.48	0.0239	2.06	0.000	0.000
UI2_BL	1314	6.70	0.54	0.0149	491	6.57	0.50	0.0226	2.06	0.000	0.000
UC_BL	1920	8.54	0.61	0.0138	644	8.19	0.55	0.0216	4.38	0.000	0.000
UP3_BL	2503	9.67	0.69	0.0137	753	9.45	0.66	0.0240	2.38	0.000	0.000
UP4_BL	2514	9.58	0.67	0.0134	747	9.30	0.63	0.0231	3.00	0.000	0.000
UM1_BL	3330	11.78	0.63	0.0110	1050	11.52	0.62	0.0192	2.28	0.000	0.000
UM2_BL	3171	11.87	0.77	0.0137	982	11.44	0.73	0.0233	3.80	0.000	0.000
UM3_BL	2046	11.41	0.92	0.0203	609	10.93	0.84	0.0342	4.36	0.000	0.000
LI1_BL	1012	5.90	0.41	0.0128	399	5.76	0.40	0.0200	2.54	0.000	0.000
LI2_BL	1260	6.30	0.41	0.0117	485	6.18	0.40	0.0184	2.05	0.000	0.000
LC_BL	1576	7.90	0.60	0.0150	539	7.45	0.52	0.0226	6.09	0.000	0.000
LP3_BL	1893	8.23	0.63	0.0144	617	7.97	0.62	0.0249	3.20	0.000	0.000
LP4_BL	1865	8.54	0.61	0.0142	616	8.32	0.57	0.0230	2.64	0.000	0.000
LM1_BL	2290	10.87	0.60	0.0125	732	10.65	0.58	0.0214	2.08	0.000	0.000
LM2_BL	2326	10.52	0.65	0.0135	745	10.25	0.65	0.0237	2.68	0.000	0.000
LM3_BL	1855	10.37	0.74	0.0172	531	10.07	0.75	0.0324	3.01	0.000	0.000

BL, buccolingual; *MD*, mesiodistal.

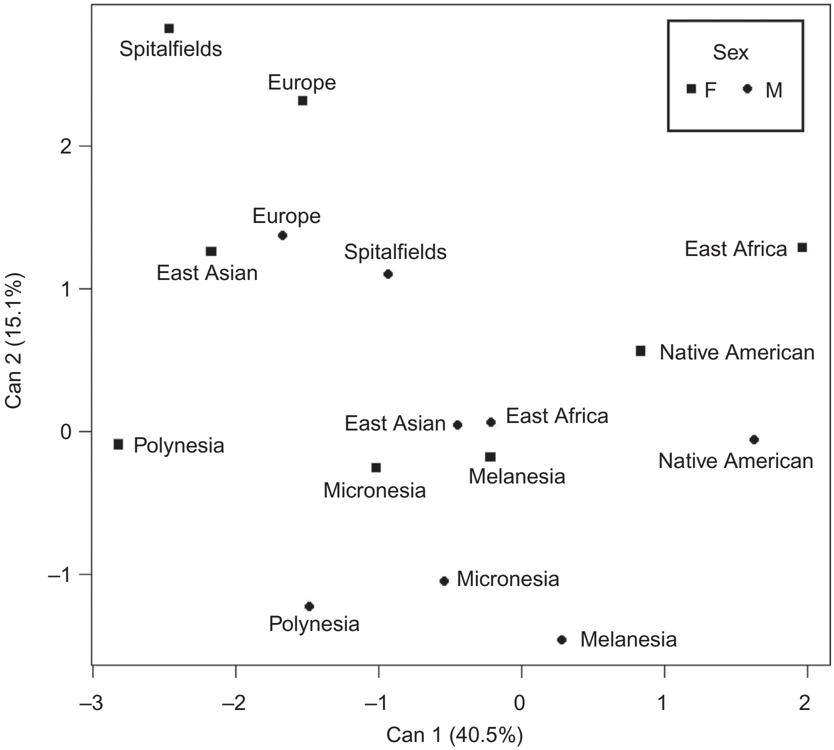


Fig. 2 Scattergram of first two canonical variates of all groups divided by sex.

however, similarities between East Asian females and European males are apparent as well as between Melanesian and Micronesian females and East Asian and African males.

Therefore, further analyses were conducted to explore sexual dimorphism by sample group. For this analysis, the broad continental groupings as listed in Table 2 are used. Summary statistics, sexual dimorphism measures, and statistical results are presented in Tables 4–6. In every sample, the buccolingual measure of the lower canine is the most dimorphic. In Asian and European samples, this is followed by the buccolingual measurement of the upper canine. However, in the African sample, the next most dimorphic measurement is the buccolingual measurement of the lower central incisor. Overall, the European sample is the most dimorphic, with the highest percentages and the most measurements showing a statistically significant difference. The Asian sample is the next most dimorphic. The African sample is the least dimorphic in terms of overall percentage and statistical differences between males and females. This result is not entirely surprising, as low levels of sexual dimorphism have also been noted for African craniometric variables (L'Abbé, Kenyhercz, Stull, Keough, & Nawrocki, 2013).

Table 4 Summary statistics for all African samples divided by sex.

	Male				Female				% Dimorphism		
	N	Mean	Std. deviation	Std. error mean	*N*	Mean	Std. deviation	Std. error mean	(M/F − 1.00)	ANOVA	*t*-test
UI1_MD	95	8.98	0.56	0.06	11	9.21	0.41	0.12	−2.53	0.19	0.11
UI2_MD	142	7.20	0.55	0.05	17	7.26	0.67	0.16	−0.71	0.72	0.76
UC_MD	305	8.00	0.49	0.03	48	7.89	0.49	0.07	1.39	0.15	0.16
UP3_MD	477	7.54	0.47	0.02	69	7.52	0.40	0.05	0.28	0.73	0.70
UP4_MD	464	7.17	0.49	0.02	68	7.11	0.45	0.05	0.77	0.39	0.36
UM1_MD	583	11.14	0.60	0.02	85	11.00	0.66	0.07	1.28	0.04	0.06
UM2_MD	579	10.58	0.75	0.03	84	10.39	0.73	0.08	1.78	0.04	0.03
UM3_MD	454	9.51	0.80	0.04	69	9.22	0.75	0.09	3.19	0.00	0.00
LI1_MD	92	5.51	0.35	0.04	16	5.50	0.29	0.07	0.04	0.98	0.98
LI2_MD	145	6.15	0.40	0.03	22	6.20	0.40	0.09	−0.85	0.57	0.58
LC_MD	225	7.31	0.49	0.03	29	7.16	0.49	0.09	2.12	0.12	0.13
LP3_MD	318	7.51	0.51	0.03	42	7.43	0.53	0.08	1.13	0.32	0.34
LP4_MD	299	7.54	0.58	0.03	44	7.56	0.48	0.07	−0.23	0.85	0.83
LM1_MD	389	11.79	0.60	0.03	47	11.74	0.46	0.07	0.40	0.61	0.53
LM2_MD	403	11.33	0.75	0.04	47	11.03	0.70	0.10	2.74	0.01	0.01
LM3_MD	364	11.25	0.91	0.05	43	10.91	0.92	0.14	3.12	0.02	0.03
UI1_BL	115	7.49	0.50	0.05	14	7.42	0.46	0.12	0.93	0.62	0.61
UI2_BL	170	6.84	0.54	0.04	24	6.67	0.44	0.09	2.64	0.13	0.08
UC_BL	319	8.69	0.60	0.03	48	8.46	0.54	0.08	2.76	0.01	0.01
UP3_BL	473	9.85	0.61	0.03	68	9.67	0.51	0.06	1.81	0.02	0.01
UP4_BL	467	9.78	0.64	0.03	68	9.63	0.57	0.07	1.52	0.08	0.06
UM1_BL	593	11.78	0.61	0.02	85	11.57	0.54	0.06	1.73	0.00	0.00
UM2_BL	588	12.04	0.79	0.03	84	11.77	0.72	0.08	2.35	0.00	0.00
UM3_BL	466	11.73	0.91	0.04	69	11.27	0.78	0.09	4.12	0.00	0.00
LI1_BL	103	5.83	0.44	0.04	17	5.55	0.30	0.07	5.08	0.01	0.00
LI2_BL	159	6.31	0.41	0.03	23	6.10	0.37	0.08	3.46	0.02	0.02
LC_BL	238	7.98	0.59	0.04	29	7.55	0.52	0.10	5.70	0.00	0.00
LP3_BL	317	8.45	0.61	0.03	42	8.22	0.47	0.07	2.72	0.02	0.01
LP4_BL	301	8.68	0.65	0.04	44	8.52	0.62	0.09	1.89	0.12	0.11
LM1_BL	383	10.93	0.59	0.03	47	10.72	0.52	0.08	1.99	0.02	0.01
LM2_BL	399	10.66	0.66	0.03	47	10.42	0.60	0.09	2.31	0.02	0.01
LM3_BL	359	10.53	0.73	0.04	42	10.16	0.64	0.10	3.69	0.00	0.00

Table 5 Summary statistics for all Asian samples divided by sex.

	Male				Female				% Dimorphism		
	N	Mean	Std. deviation	Std. error mean	*N*	Mean	Std. deviation	Std. error mean	(M/F − 1.00)	ANOVA	*t*-test
UI1_MD	725	8.72	0.58	0.02	318	8.57	0.55	0.03	1.65	0.00	0.00
UI2_MD	862	7.25	0.64	0.02	391	7.17	0.57	0.03	1.09	0.04	0.03
UC_MD	1248	8.09	0.50	0.01	509	7.89	0.47	0.02	2.63	0.00	0.00
UP3_MD	1607	7.40	0.46	0.01	579	7.24	0.44	0.02	2.16	0.00	0.00
UP4_MD	1599	7.03	0.50	0.01	568	6.92	0.47	0.02	1.57	0.00	0.00
UM1_MD	2182	10.91	0.62	0.01	823	10.73	0.61	0.02	1.73	0.00	0.00
UM2_MD	2015	10.29	0.70	0.02	757	10.07	0.68	0.02	2.18	0.00	0.00
UM3_MD	1257	9.36	0.85	0.02	452	9.34	0.79	0.04	0.26	0.60	0.59
LI1_MD	636	5.49	0.35	0.01	296	5.50	0.34	0.02	−0.29	0.52	0.51
LI2_MD	798	6.16	0.43	0.02	363	6.19	0.41	0.02	−0.43	0.32	0.31
LC_MD	997	7.16	0.47	0.02	415	6.90	0.44	0.02	3.83	0.00	0.00
LP3_MD	1242	7.25	0.49	0.01	488	7.08	0.43	0.02	2.37	0.00	0.00
LP4_MD	1238	7.34	0.54	0.02	480	7.23	0.50	0.02	1.56	0.00	0.00
LM1_MD	1530	11.71	0.62	0.02	587	11.49	0.61	0.03	1.91	0.00	0.00
LM2_MD	1528	11.19	0.80	0.02	605	10.94	0.75	0.03	2.30	0.00	0.00
LM3_MD	1190	11.21	0.94	0.03	424	10.79	0.98	0.05	3.86	0.00	0.00
UI1_BL	813	7.41	0.49	0.02	334	7.28	0.47	0.03	1.81	0.00	0.00
UI2_BL	927	6.74	0.53	0.02	395	6.62	0.46	0.02	1.75	0.00	0.00
UC_BL	1278	8.56	0.60	0.02	506	8.23	0.51	0.02	3.99	0.00	0.00
UP3_BL	1603	9.82	0.59	0.01	574	9.58	0.57	0.02	2.49	0.00	0.00
UP4_BL	1596	9.65	0.64	0.02	560	9.37	0.58	0.02	3.00	0.00	0.00
UM1_BL	2206	11.87	0.62	0.01	822	11.60	0.60	0.02	2.36	0.00	0.00
UM2_BL	2018	11.92	0.74	0.02	752	11.50	0.70	0.03	3.68	0.00	0.00
UM3_BL	1252	11.39	0.88	0.02	457	10.98	0.83	0.04	3.75	0.00	0.00
LI1_BL	704	5.93	0.40	0.02	315	5.80	0.40	0.02	2.32	0.00	0.00
LI2_BL	845	6.33	0.41	0.01	378	6.21	0.40	0.02	1.85	0.00	0.00
LC_BL	1037	7.92	0.60	0.02	423	7.50	0.50	0.02	5.61	0.00	0.00
LP3_BL	1236	8.32	0.57	0.02	484	8.08	0.55	0.03	3.01	0.00	0.00
LP4_BL	1224	8.60	0.58	0.02	475	8.39	0.54	0.02	2.55	0.00	0.00
LM1_BL	1527	10.96	0.57	0.01	592	10.73	0.55	0.02	2.18	0.00	0.00
LM2_BL	1518	10.59	0.63	0.02	600	10.32	0.62	0.03	2.65	0.00	0.00
LM3_BL	1186	10.44	0.71	0.02	425	10.14	0.74	0.04	3.02	0.00	0.00

Table 6 Summary statistics for all European samples divided by sex.

	Male				Female				% Dimorphism		
	N	Mean	Std. deviation	Std. error mean	*N*	Mean	Std. deviation	Std. error mean	(M/F − 1.00)	ANOVA	*t*-test
UI1_MD	172	8.43	0.50	0.04	49	8.22	0.47	0.07	2.62	0.008	0.007
UI2_MD	195	6.61	0.50	0.04	66	6.49	0.56	0.07	1.87	0.102	0.124
UC_MD	314	7.61	0.45	0.03	91	7.34	0.41	0.04	3.65	0.000	0.000
UP3_MD	428	6.76	0.42	0.02	111	6.57	0.43	0.04	2.88	0.000	0.000
UP4_MD	445	6.55	0.42	0.02	118	6.44	0.45	0.04	1.65	0.015	0.021
UM1_MD	527	10.49	0.53	0.02	143	10.21	0.65	0.05	2.79	0.000	0.000
UM2_MD	559	9.77	0.60	0.03	146	9.41	0.60	0.05	3.87	0.000	0.000
UM3_MD	329	9.00	0.71	0.04	83	8.72	0.66	0.07	3.13	0.002	0.001
LI1_MD	180	5.22	0.29	0.02	58	5.12	0.33	0.04	1.86	0.039	0.054
LI2_MD	234	5.83	0.36	0.02	81	5.67	0.39	0.04	2.75	0.001	0.002
LC_MD	286	6.69	0.45	0.03	79	6.39	0.36	0.04	4.78	0.000	0.000
LP3_MD	338	6.72	0.46	0.02	91	6.47	0.46	0.05	3.94	0.000	0.000
LP4_MD	338	6.92	0.46	0.02	100	6.77	0.51	0.05	2.30	0.004	0.007
LM1_MD	362	11.09	0.57	0.03	88	10.73	0.66	0.07	3.35	0.000	0.000
LM2_MD	408	10.70	0.69	0.03	98	10.29	0.64	0.07	4.05	0.000	0.000
LM3_MD	309	10.69	0.87	0.05	66	10.26	0.73	0.09	4.14	0.000	0.000
UI1_BL	195	7.20	0.47	0.03	56	6.92	0.45	0.06	4.04	0.000	0.000
UI2_BL	217	6.46	0.52	0.04	72	6.25	0.63	0.07	3.32	0.006	0.013
UC_BL	323	8.34	0.58	0.03	90	7.79	0.55	0.06	7.10	0.000	0.000
UP3_BL	427	8.92	0.59	0.03	111	8.61	0.55	0.05	3.57	0.000	0.000
UP4_BL	451	9.12	0.59	0.03	119	8.78	0.61	0.06	3.83	0.000	0.000
UM1_BL	531	11.42	0.58	0.03	143	11.04	0.59	0.05	3.46	0.000	0.000
UM2_BL	565	11.50	0.73	0.03	146	10.92	0.65	0.05	5.32	0.000	0.000
UM3_BL	328	11.04	0.93	0.05	83	10.41	0.78	0.09	6.02	0.000	0.000
LI1_BL	205	5.85	0.40	0.03	67	5.63	0.38	0.05	3.88	0.000	0.000
LI2_BL	256	6.21	0.43	0.03	84	6.02	0.38	0.04	3.06	0.000	0.000
LC_BL	301	7.75	0.59	0.03	87	7.13	0.54	0.06	8.63	0.000	0.000
LP3_BL	340	7.69	0.53	0.03	91	7.30	0.59	0.06	5.30	0.000	0.000
LP4_BL	340	8.19	0.55	0.03	97	7.90	0.53	0.05	3.70	0.000	0.000
LM1_BL	380	10.44	0.50	0.03	93	10.11	0.46	0.05	3.31	0.000	0.000
LM2_BL	409	10.12	0.57	0.03	98	9.72	0.56	0.06	4.13	0.000	0.000
LM3_BL	310	9.91	0.70	0.04	64	9.55	0.67	0.08	3.79	0.000	0.000

Conclusions

While tooth crown measurements help distinguish males from females, this is not true for crown and root morphology. The presence or absence of a dental morphological trait is of no utility in sex estimation. In a forensic context, even if an individual exhibited a pronounced distal accessory ridge on the lower canine, such a finding would only slightly increase the odds that the individual was male. If an individual has teeth, only tooth size—not morphology—is a useful adjunct to other methods of sex estimation.

In the formulation of methods to estimate sex from odontometrics, it is critical to consider population variation and use the most appropriate method given the set of circumstances to avoid misclassifications. Further, models should incorporate the upper and lower canine teeth if possible. While data on third molars are presented here, it is generally advisable to omit these teeth from such methods. These teeth are highly variable and may often be missing congenitally or through antemortem extraction. Univariate methods are discouraged, as multivariate methods perform better in odontometric sex estimations (Martins Filho et al., 2016).

The odontometric data collected by Hanihara, published in Pilloud et al. (2014), and described herein can also be used as a custom dataset in the statistical application FORDISC (Jantz & Ousley, 2005), using the custom import function (see Chapter 12 of this volume). In this way it is possible to estimate sex using dental metrics in much the same way as cranial metrics (Fig. 3). This dataset can be requested from the lead author

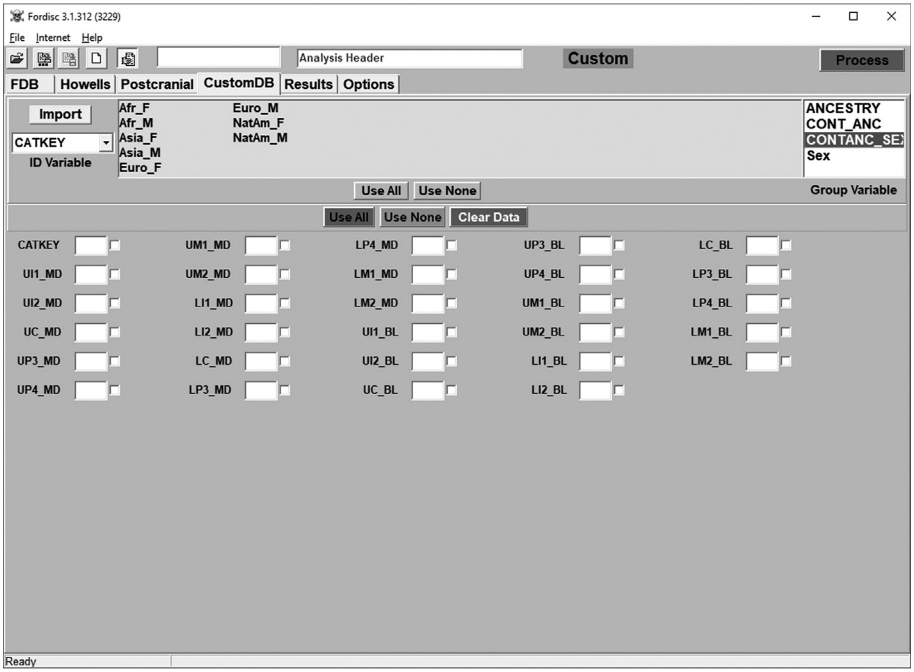


Fig. 3 Example of odontometric data for use in FORDISC.

for use in forensic anthropological casework. Both authors are also actively collecting data on crown and cervical metrics on modern samples to grow the application of these data. A web-based application is currently under development to aid in these efforts.

Acknowledgments

We thank Tsunehiko Hanihara for generously sharing his data. We also thank Casey Philbin for statistical consultation.

References

Aas, I. H. (1979). The depth of the lingual fossa in permanent maxillary incisors of Norwegian Lapps. *American Journal of Physical Anthropology*, *51*(3), 417–419.

Abrantes, C., Santos, R., Pestana, D., & Pereira, C. P. (2015). Application of dental morphological characteristics for medical-legal identification: Sexual diagnosis in a Portuguese population. *HSOA Journal of Forensic, Legal & Investigative Sciences*, *1*, 001.

Acharya, A. B., Angadi, P. V., Prabhu, S., & Nagnur, S. (2011). Validity of the mandibular canine index (MCI) in sex prediction: Reassessment in an Indian sample. *Forensic Science International*, *204*(1), 207.e201–207.e204.

Acharya, A. B., & Mainali, S. (2007). Univariate sex dimorphism in the Nepalese dentition and the use of discriminant functions in gender assessment. *Forensic Science International*, *173*(1), 47–56.

Acharya, A. B., & Mainali, S. (2009). Limitations of the mandibular canine index in sex assessment. *Journal of Forensic and Legal Medicine*, *16*(2), 67–69.

Adams, D. M., & Pilloud, M. A. (2019). Sex estimation from dental crown and cervical metrics in a contemporary Japanese sample. *Forensic Anthropology*, *2*(4), 1–11.

Adler, C. J., & Donlon, D. (2010). Sexual dimorphism in deciduous crown traits of a European derived Australian sample. *Forensic Science International*, *199*(1–3), 29–37.

Alvesalo, L. (1997). Sex chromosomes and human growth: A dental approach. *Human Genetics*, *101*(1–5).

Alvesalo, L. (2009). Human sex chromosomes in oral and craniofacial growth. *Archives of Oral Biology*, *54*, S18–S24.

Alvesalo, L. (2013). The expression of human sex chromosome genes in oral and craniofacial growth. In G. R. Scott & J. D. Irish (Eds.), *Anthropological perspectives on tooth morphology: Genetics, evolution, variation* (pp. 92–107). Cambridge: Cambridge University Press.

Alvesalo, L., & Laine, T. (1992). Occlusion in 47,XXY (Klinefelter syndrome) men. *The American Journal of Physical Anthropology*, *87*(2), 161–165.

Alvesalo, L., Tammisalo, E., & Townsend, G. (1991). Upper central incisor and canine tooth crown size in 47,XXY males. *Journal of Dental Research*, *70*(7), 1057–1060.

Angadi, P. V., Hemani, S., Prabhu, S., & Acharya, A. B. (2013). Analyses of odontometric sexual dimorphism and sex assessment accuracy on a large sample. *Journal of Forensic and Legal Medicine*, *20*(6), 673–677.

Anuthama, K., Shankar, S., Ilayaraja, V., Kumar, G. S., Rajmohan, M., & Vignesh. (2011). Determining dental sex dimorphism in South Indians using discriminant function analysis. *Forensic Science International*, *212*(1), 86–89.

Bakkannavar, S. M., Manjunath, S., Nayak, V. C., & Pradeep Kumar, G. (2015). Canine index – A tool for sex determination. *Egyptian Journal of Forensic Sciences*, *5*(4), 157–161.

Bang, G., & Hasund, A. (1972). Morphologic characteristics of the Alaskan Eskimo dentition. II. Carabelli's cusp. *The American Journal of Physical Anthropology*, *37*(1), 35–39.

Black, T. K. (1978). Sexual dimorphism in the tooth-crown diameters of the deciduous teeth. *The American Journal of Physical Anthropology*, *48*(1), 77–82.

Brace, C. L., & Ryan, A. S. (1980). Sexual dimorphism and human tooth size differences. *Journal of Human Evolution*, *9*(5), 417–435.

Brook, A., Griffin, R., Townsend, G., Levisianos, Y., Russell, J., & Smith, R. (2009). Variability and patterning in permanent tooth size of four human ethnic groups. *Archives of Oral Biology*, *54*, S79–S85.

Calhoun, G. V., Guatelli-Steinberg, D., & Hubbe, M. (2018). *Quantifying the sexual dimorphism in the geometry of modern human canines*. Austin, Texas: American Association of Physical Anthropologists.

Cardoso, H. F. V. (2010). Testing discriminant functions for sex determination from deciduous teeth. *Journal of Forensic Sciences*, *55*(6), 1557–1560.

Chan, A. H. (2007). *Sexual dimorphism of modern tooth crown tissues*. Anthropology McMaster University.

Coppa, A., Cucina, A., Mancinelli, D., Vargiu, R., & Calcagno, J. M. (1998). Dental anthropology of central-southern, iron age Italy: The evidence of metric versus nonmetric traits. *The American Journal of Physical Anthropology*, *107*, 371–386.

Cucina, A., Lucci, M., Vargiu, R., & Coppa, A. (1999). Dental evidence of biological affinity and environmental conditions in prehistoric Trentino (Italy) samples from the Neolithic to the Early Bronze Age. *International Journal of Osteoarchaeology*, *9*(6), 404–416.

De Angelis, D., Gibelli, D., Gaudio, D., Cipriani Noce, F., Guercini, N., Varvara, G., et al. (2015). Sexual dimorphism of canine volume: A pilot study. *Legal Medicine*, *17*(3), 163–166.

DeVito, C., & Saunders, S. R. (1990). A discriminant function analysis of deciduous teeth to determine sex. *Journal of Forensic Sciences*, *35*(4), 845–858.

El Sheikhi, F., & Bugaighis, I. (2016). Sex discrimination by odontometrics in Libyan subjects. *Egyptian Journal of Forensic Sciences*, *6*(2), 157–164.

Garcia-Campos, C., Martinon-Torres, M., Martin-Frances, L., Martinez de Pinillos, M., Modesto-Mata, M., Perea-Perez, B., et al. (2018). Contribution of dental tissues to sex determination in modern human populations. *American Journal of Physical Anthropology*, *166*(2), 459–472.

Garn, S., Kerewsky, R. S., & Lewis, A. B. (1966). Extent of sex influence on Carabelli's polymorphism. *Journal of Dental Research*, *45*, 1823.

Garn, S. M., Cole, P. E., Wainwright, R. L., & Guire, K. E. (1977). Sex discriminatory effectiveness using combinations of permanent teeth. *Journal of Dental Research*, *56*, 697.

Garn, S. M., Kerewsky, R. S., & Swindler, D. R. (1966). Canine "field" in sexual dimorphism of tooth size. *Nature*, *212*, 1501.

Garn, S. M., Lewis, A. B., & Kerewsky, R. S. (1964). Sex difference in tooth size. *Journal of Dental Research*, *43*, 306.

Garn, S. M., Lewis, A. B., Swindler, D. R., & Kerewsky, R. S. (1967). Genetic control of sexual dimorphism in tooth size. *Journal of Dental Research*, *46*(5), 963–973.

Gingerich, P. D. (1974). Size variability o the teeth in living mammals and the diagnosis of closely related sympatric fossil species. *Journal of Paleontology*, *48*, 895–903.

Goose, D. H., & Lee, G. (1971). The mode of inheritance of Carabelli's trait. *Human Biology*, *43*(1), 64–69.

Guatelli-Steinberg, D., Sciulli, P. W., & Betsinger, T. K. (2008). Dental crown size and sex hormone concentrations: Another look at the development of sexual dimorphism. *American Journal of Physical Anthropology*, *137*(3), 324–333.

Hanihara, T. (1998). Metric and nonmetric dental variations of major human populations. In J. R. Lukacs (Ed.), *Human dental development, morphology, and pathology: A tribute to Albert A. Dahlberg* (pp. 173–200). Eugene, OR: Department of Anthropology, University of Oregon.

Hanihara, T., & Ishida, H. (2005). Metric dental variation of major human populations. *The American Journal of Physical Anthropology*, *128*, 287–298.

Harris, E. F. (1980). Sex differences in lingual marginal ridging on the human maxillary central incisor. *The American Journal of Physical Anthropology*, *52*, 541–548.

Harris, E. F., & Couch, W. M. J. (2006). The relative sexual dimorphism of human incisor crown and root dimensions. *Dental Anthropology*, *19*(3), 87–95.

Harris, E. F., & Dinh, D. P. (2006). Intercusp relationships of the permanent maxillary first and second molars in American whites. *The American Journal of Physical Anthropology*, *130*(4), 514–528.

Harris, E. F., & Foster, C. L. (2015). Size matters: Discrimination between American blacks and whites, males and females, using tooth crown dimensions. In G. E. Berg & S. C. Ta'ala (Eds.), *Biological affinity in forensic identification of human skeletal remains* (pp. 209–238). Boca Raton, FL: CRC Press.

Harris, E. F., Hicks, J. D., & Barcroft, B. D. (2001). Tissue contributions to sex and race: Differences in tooth crown size of deciduous molars. *The American Journal of Physical Anthropology*, *115*(3), 223–237.

Harris, E. F., & Lease, L. R. (2005). Mesiodistal tooth crown dimensions of the primary dentition: A worldwide survey. *The American Journal of Physical Anthropology*, *128*, 593–607.

Hassett, B. (2012). Technical note: Estimating sex using cervical canine odontometrics: A test using a known sex sample. *The American Journal of Physical Anthropology*, *146*, 486–489.

Heim, K. (2018). *Population variation in dental development and its effect on forensic age estimation*. [unpublished PhD dissertation] Anthropology Reno: University of Nevada.

Hillson, S., FitzGerald, C., & Flinn, H. (2005). Alternative dental measurements: Proposals and relationships with other measurements. *The American Journal of Physical Anthropology*, *126*, 413–426.

Humphrey, L. T. (1998). Growth patterns in the modern human skeleton. *American Journal of Physical Anthropology*, *105*(1), 57–72.

Iqbal, R., Zhang, S., & Mi, C. (2015). Reliability of mandibular canine and mandibular canine index in sex determination: A study using Uyghur population. *Journal of Forensic and Legal Medicine*, *33*, 9–13.

Irish, J. D. (2006). Who were the ancient Egyptians? Dental affinities among Neolithic through postdynastic peoples. *The American Journal of Physical Anthropology*, *129*, 529–543.

İşcan, M. Y., & Kedici, P. S. (2003). Sexual variation in bucco-lingual dimensions in Turkish dentition. *Forensic Science International*, *137*(2), 160–164.

Jain, A. K., Garg, N., Singh, J., Ansari, A., & Sangamesh, B. (2011). Mesiodistal crown dimensions of the permanent dentition of a North Indian population. *Indian Journal of Dentistry*, *2*(2), 16–20.

Jantz, R., & Ousley, S. (2005). *FORDISC 3: Computerized forensic discriminant functions*. Knoxville: The University of Tennessee.

Joshi, M., Godiawala, R., & Dutia, A. (1972). Carabelli's trait in Hindu children from Gujarat. *Journal of Dental Research*, *51*(3), 706–711.

Kaul, V., & Prakash, S. (1981). Morphological features of Jat dentition. *The American Journal of Physical Anthropology*, *54*(1), 123–127.

Kaur, S., & Chattopadhyay, P. K. (2003). Sexual dimorphism of incisors: A study of the Jat Sikhs. *Legal Medicine*, *5*, S261–S262.

Kazzazi, S. M., & Kranioti, E. F. (2018). Sex estimation using cervical dental measurements in an archaeological population from Iran. *Archaeological and Anthropological Sciences*, *10*(2), 439–448.

Khamis, M. F., Taylor, J. A., Malik, S. N., & Townsend, G. C. (2014). Odontometric sex variation in Malaysians with application to sex prediction. *Forensic Science International*, *234*, 183.e181–183.e187.

Kieser, J., & Preston, C. (1981). The dentition of the Lengua Indians of Paraguay. *The American Journal of Physical Anthropology*, *55*, 485–490.

Kimbel, W. H., & Delezene, L. K. (2009). "Lucy" redux: A review of research on *Australopithecus afarensis*. *The American Journal of Physical Anthropology*, *52*, 2–48.

Kimura, R., Yamaguchi, T., Takeda, Y., Kondo, O., Toma, T., Haneji, K., et al. (2009). A common variation in EDAR is a genetic determinant of shovel-shaped incisors. *The American Journal of Human Genetics*, *84*(4), 528–535.

Kirveskari, P., & Alvesalo, L. (1981). Shovel shape of maxillary incisors in 47, XXY males. *Proceedings of the Finnish Dental Society Suomen Hammaslaakariseuran toimituksia*, 77(1–3), 79–81.

Kirveskari, P., & Alvesalo, L. (1982). Dental morphology in Turner's syndrome (45, X females). In B. Kurtén (Ed.), *Teeth: Form, function, and evolution* (pp. 298–303). New York: Columbia University.

Kondo, S., & Townsend, G. C. (2004). Sexual dimorphism in crown units of mandibular deciduous and permanent molars in Australian aborigines. *HOMO – Journal of Comparative Human Biology*, *55*(1), 53–64.

Kondo, S., Townsend, G. C., & Yamada, H. (2005). Sexual dimorphism of cusp dimensions in human maxillary molars. *The American Journal of Physical Anthropology*, *128*(4), 870–877.

Kondoh, Y. (1990). A study on the dental traits and crown sizes of Japanese children. *Aichi Gakuin Daigaku Shigakkai Shi*, *28*(1 Pt 1), 159–194.

L'Abbé, E. N., Kenyhercz, M., Stull, K. E., Keough, N., & Nawrocki, S. (2013). Application of Fordisc 3.0 to explore differences among crania of North American and South African blacks and whites. *Journal of Forensic Sciences*, *58*(6), 1579–1583.

Lähdesmäki, R., & Alvesalo, L. (2006). Root growth in the permanent teeth of 45,X/46,XX females. *European Journal of Orthodontics*, *28*(4), 339–344.

Lähdesmäki, R., & Alvesalo, L. (2007). Root lengths in the permanent teeth of Klinefelter (47, XXY) men. *Archives of Oral Biology*, *52*(9), 822–827.

Laine, T., Alvesalo, L., & Lammi, S. (1992). A study in 47,XYY men of the expression of sex-chromosome anomalies in dental occlusion. *Archives of Oral Biology*, *37*(11), 923–928.

Lau, E. C., Mohandas, T. K., Shapiro, L. J., Slavkin, H. C., & Snead, M. L. (1989). Human and mouse amelogenin gene loci are on the sex chromosomes. *Genomics*, *4*(2), 162–168.

Leutenegger, W., & Shell, B. (1987). Variability and sexual dimorphism in canine size of Australopithecus and extant hominoids. *Journal of Human Evolution*, *16*(4), 359–367.

Macaluso, P. J. (2011). Investigation on the utility of permanent maxillary molar cusp areas for sex estimation. *Forensic Science, Medicine, and Pathology*, 7(3), 233–247.

Manabe, Y., Oyamada, J., Kitagawa, Y., Rokutanda, A., Kato, K., & Matsushita, T. (2003). Dental morphology of the Dawenkou Neolithic population in North China: Implications for the origin and distribution of sinodonty. *Journal of Human Evolution*, *45*(5), 369–380.

Manchanda, A. S., Narang, R. S., Kahlon, S. S., & Singh, B. (2015). Diagonal tooth measurements in sex assessment: A study on North Indian population. *Journal of Forensic Dental Sciences*, 7(2), 126–131.

Martins Filho, I. E., Lopez-Capp, T. T., Biazevic, M. G. H., & Michel-Crosato, E. (2016). Sexual dimorphism using odontometric indexes: Analysis of three statistical techniques. *Journal of Forensic and Legal Medicine*, *44*, 37–42.

Matsumura, H. (2007). Non-metric dental trait variation among local sites and regional groups of the Neolithic Jomon period, Japan. *Anthropological Science*, *115*(1), 25–33.

Mayhall, J. T. (1992). Techniques for the study of dental morphology. In S. R. Saunders & M. A. Katzenberg (Eds.), *The skeletal biology of past peoples: Research methods* (pp. 59–78). New York: Wiley-Liss.

Mayhall, J. T., Alvesalo, L., & Townsend, G. C. (1998). Dental morphology of 47,XYY males: molar cusp area, volume, shape, and linear dimensions. In J. R. Lukacs (Ed.), *Human dental development, morphology and pathology: A tribute to Albert A. Dahlberg* (pp. 29–39). Eugene: University of Oregon.

Mitsea, A. G., Moraitis, K., Leon, G., Nicopoulou-Karayianni, K., & Spiliopoulou, C. (2014). Sex determination by tooth size in a sample of Greek population. *HOMO – Journal of Comparative Human Biology*, *65*(4), 322–329.

Mizoguchi, Y. (1985). *Shovelling: A statistical analysis of its morphology*. Tokyo: University of Tokyo Press.

Moorrees, C. F. A. (1957). *The Aleut dentition*. Cambridge, Massachusetts: Harvard University Press.

Moorrees, C. F. A., & Reed, R. B. (1964). Correlations among crown diameters of human teeth. *Archives of Oral Biology*, *9*, 685–697.

Nanci, A. (Ed.), (2003). *Ten Cate's oral histology: Development, structure, and function*. St. Louis, Missouri: Mosby.

Noss, J. F., Scott, R. G., Potter, R. H. Y., Dahlberg, A., & Dahlberg, T. (1983). The influence of crown size dimorphism on sex differences in the Carabelli trait and the canine distal accessory ridge in man. *Archives of Oral Biology*, *28*, 527–530.

Park, J. -H., Yamaguchi, T., Watanabe, C., Kawaguchi, A., Haneji, K., Takeda, M., et al. (2012). Effects of an Asian-specific nonsynonymous EDAR variant on multiple dental traits. *Journal of Human Genetics*, *57*(8), 508–514.

Peckmann, T. R., Logar, C., Garrido-Varas, C. E., Meek, S., & Pinto, X. T. (2016). Sex determination using the mesio-distal dimension of permanent maxillary incisors and canines in a modern Chilean population. *Science & Justice*, *56*(2), 84–89.

Peckmann, T. R., Meek, S., Dilkie, N., & Mussett, M. (2015). Sex estimation using diagonal diameter measurements of molar teeth in African American populations. *Journal of Forensic and Legal Medicine*, *36*, 70–80.

Pedersen, P. O. (1949). *The East Greenland Eskimo dentition, numerical variations and anatomy*. Copenhagen, Denmark: CA Reitzel.

Pereira, C., Bernardo, M., Pestana, D., Santos, J. C., & de Mendonça, M. C. (2010). Contribution of teeth in human forensic identification – Discriminant function sexing odontometrical techniques in Portuguese population. *Journal of Forensic and Legal Medicine*, *17*(2), 105–110.

Pilloud, M. A., Hefner, J. T., Hanihara, T., & Hayashi, A. (2014). The use of tooth crown measurements in the assessment of ancestry. *Journal of Forensic Sciences*, *59*(6), 1493–1501.

Pilloud, M. A., & Hillson, S. (2012). Brief communication: The use of alternative dental measurements on deciduous teeth. *The American Journal of Physical Anthropology*, *149*, 299–306.

Pinhasi, R., & Meiklejohn, C. (2011). Dental reduction and the transition to agriculture in Europe. In R. Pinhasi & J. T. Stock (Eds.), *Human bioarchaeology of the transition to agriculture* (pp. 451–474): John Wiley & Sons Ltd.

Plavcan, J. M. (2001). Sexual dimorphism in primate evolution. *The American Journal of Physical Anthropology*, *44*, 25–53.

Plavcan, J. M. (2012). Sexual size dimorphism, canine dimorphism, and male-male competition in primates. *Human Nature*, *23*(1), 45–67.

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, ISBN 3-900051-07-0. http://www.R-project.org.

Ribeiro, D. C., Brook, A. H., Hughes, T. E., Sampson, W. J., & Townsend, G. C. (2013). Intrauterine hormone effects on tooth dimensions. *Journal of Dental Research*, *92*(5), 425–431.

Rothhammer, F., Lasserre, E., Blanco, R., Covarrubias, E., & Dixon, M. (1968). Microevolution in human Chilean populations: IV. Shovel shape, mesial-palatal version and other dental traits in Pewenche Indians. *Zeitschrift für Morphologie und Anthropologie*, *60*, 162–169.

Saikiran, C., Khaitan, T., Ramaswamy, P., Sudhakar, S., Smitha, B., & Uday, G. (2014). Role of mandibular canines in establishment of gender. *Egyptian Journal of Forensic Sciences*, *4*(3), 71–74.

Saunders, S. R., Chan, A. H., Kahlon, B., Kluge, H. F., & FitzGerald, C. M. (2007). Sexual dimorphism of the dental tissues in human permanent mandibular canines and third premolars. *The American Journal of Physical Anthropology*, *133*, 735–740.

Scheid, R. C., & Weiss, G. (2012). *Woelfel's dental anatomy*. Philadelphia, PA: Lippincot, Williams, & Wilkins.

Schnutenhaus, S., & Rösing, F. W. (1998). World variation of tooth size. In K. W. Alt, F. W. Rösing, & M. Teschler-Nicola (Eds.), *Dental anthropology: Fundamentals, limits, and prospects* (pp. 522–535). New York: Springer.

Schulman, G. S., Redford-Badwal, D., Poole, A., Mathieu, G., Burleson, J., & Dauser, D. (2005). Taurodontism and learning disabilities in patients with Klinefelter syndrome. *Pediatric Dentistry*, *27*(5), 389–394.

Schwartz, G. T., & Dean, C. M. (2001). Ontogeny of canine dimorphism in extant hominoids. *The American Journal of Physical Anthropology*, *115*(3), 269–283.

Schwartz, G. T., & Dean, M. C. (2005). Sexual dimorphism in modern human permanent teeth. *The American Journal of Physical Anthropology*, *128*(2), 312–317.

Scott, G. R. (1977). Classification, sex dimorphism, association, and population variation of the canine distal accessory ridge. *Human Biology*, *49*(3), 453–469.

Scott, G. R. (1980). Population variation of Carabelli's trait. *Human Biology*, *52*, 63–78.

Scott, G. R., Anta, A., Schomberg, R., & De La Rue, C. (2013). Basque dental morphology and the "Eurodont" dental pattern. In G. R. Scott & J. D. Irish (Eds.), *Anthropological perspectives on tooth morphology: Genetics, evolution, variation* (pp. 296–318). Cambridge: Cambridge University Press.

Scott, G. R., & Irish, J. D. (2017). *Human crown and root morphology: The Arizona State University dental anthropology system*. Cambridge: Cambridge University Press.

Scott, G. R., Potter, R. H., Noss, J. F., Dahlberg, A. A., & Dahlberg, T. (1983). The dental morphology of Pima Indians. *The American Journal of Physical Anthropology*, *61*(1), 13–31.

Scott, G. R., & Turner, C. G. (1997). *The anthropology of modern human teeth*. Cambridge: Cambridge University Press.

Shaweesh, A. I. (2017). Mesiodistal and faciolingual diameters of the permanent teeth in a Jordanian population. *Archives of Oral Biology*, *73*, 253–258.

Silva, A. M., Pereira, M. L., Gouveia, S., Tavares, J. N., Azevedo, Á., & Caldas, I. M. (2016). A new approach to sex estimation using the mandibular canine index. *Medicine, Science, and the Law*, *56*(1), 7–12.

Smith, B. H. (1994). Patterns of dental development in Homo, Australopithecus, pan, and gorilla. *American Journal of Physical Anthropology*, *94*(3), 307–325.

Smith, T. M., Olejniczak, A. J., Reid, D. J., Ferrell, R. J., & Hublin, J. J. (2006). Modern human molar enamel thickness and enamel-dentine junction shape. *Archives of Oral Biology*, *51*(11), 974–995.

SPSS Inc. (2016). *IBM SPSS statistics for windows*. Verson 24.0 Armonk, NY: IBM Corp.

Swindler, D. R. (2002). *Primate dentition: An introduction to the teeth of non-human primates*. Cambridge: Cambridge University Press.

Townsend, G., & Brown, T. (1981). The Carabelli trait in Australian aboriginal dentition. *Archives of Oral Biology*, *26*(10), 809–814.

Townsend, G. C. (1992). Genetic and environmental contributions to morphometric dental variation. In J. R. Lukacs (Ed.), *Culture, ecology and dental anthropology* (pp. 61–72). Delhi, India: Kamla-Raj Enterprises.

Turner, C. G. (1969). Microevolutionary interpretations from the dentition. *The American Journal of Physical Anthropology*, *30*(3), 421–426.

Turner, C. G., Nichol, C. R., & Scott, G. R. (1991). Scoring procedures for key morphological traits of the permanent dentition: The Arizona State University dental anthropology system. In M. A. Kelley & C. S. Larsen (Eds.), *Advances in dental anthropology* (pp. 13–31). New York: Wiley-Liss.

Tuttösí, P., & Cardoso, H. F. V. (2015). An assessment of sexual dimorphism and sex estimation using cervical dental measurements in a Northwest Coast archeological sample. *Journal of Archaeological Science: Reports*, *3*, 306–312.

Varrela, J. (1992). Multirooted mandibular premolars in 45, X females: Frequency and morphological types. In P. Smith & E. Tchernov (Eds.), *Structure, function and evolution of teeth* (pp. 519–526). London: Freund Publishing House Ltd.

Varrela, J., & Alvesalo, L. (1988). Taurodontism in 47,XXY males: An effect of the extra X chromosome on root development. *Journal of Dental Research*, *67*(2), 501–502.

Viciano, J., López-Lázaro, S., & Alemán, I. (2013). Sex estimation based on deciduous and permanent dentition in a contemporary Spanish population. *The American Journal of Physical Anthropology*, *152*(1), 31–43.

Zorba, E., Moraitis, K., & Manolis, S. (2011). Sexual dimorphism in permanent teeth of modern Greeks. *Forensic Science International*, *210*, 74–81.

Zorba, E., Vanna, V., & Moraitis, K. (2014). Sexual dimorphism of root length on a Greek population sample. *HOMO – Journal of Comparative Human Biology*, *65*(2), 143–154.

CHAPTER 11

Metric methods for estimating sex utilizing the pelvis

Sarah E. Baumgarten[a], **Brittany Kenyon-Flatt**[b]
[a]Anthropology, Washington University in St. Louis, St. Louis, MO, United States
[b]Anthropology, University of Buffalo, Buffalo, NY, United States

Historical overview of metric methods

The initial metric methods: Indices

The pelvis has been long considered by anthropologists as the best indicator of sex for skeletal remains, largely due to the high levels of sexual dimorphism that is more predominately present in the pelvis than in other regions of the body. This dimorphism becomes apparent after puberty when, under the influence of hormones, the pelvis undergoes morphological changes, mostly concerning the bony morphology of the birth canal (Bass, 2005; Byers, 2002; Gonzalez, Bernal, & Perez, 2009; Mestekova, Bruzek, Veleminska, & Chaumoitre, 2015; Pickering & Bachman, 1997; Spradley & Jantz, 2011).

An early example of the anthropological use of innominate measurements to estimate sex in skeletal remains began with the ischio-pubic index, which was modified by Washburn in 1949 from a model previously published for use on primate skeletal remains by Schultz (1930). The ischio-pubic index can be calculated by dividing the length of the pubis by the length of the ischium, multiplied by 100. Both the length of the pubis and the length of the ischium are measured from a landmark known as the acetabular point, which is defined as the point in the acetabulum where the ilium, ischium, and pubis fuse together (Fig. 1). However, in most adult individuals, this point is extremely difficult to locate as the line of fusion is obliterated, which can dramatically increase both intra- and interobserver error rates and reliability. Various attempts have been made to remedy this definition to increase its utility, with varying success (Brauer, 1988; Gaillard, 1961; Genoves, 1959; Moeschler, 1964). Washburn (1949) additionally examined sciatic notch width in the Bantu and Bushmen populations. Through the utilization of both the ischio-pubic index and sciatic notch width, Washburn purported a correct classification accuracy of 98% for the Bantu and Bushmen populations.

In 1957, Thieme and Schull followed in the study of the ischio-pubic index with a study comprising 200 individuals from the Robert J. Terry skeletal collection. When using the ischio-pubic index solely as a means of sorting females and males, a classification

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00011-0

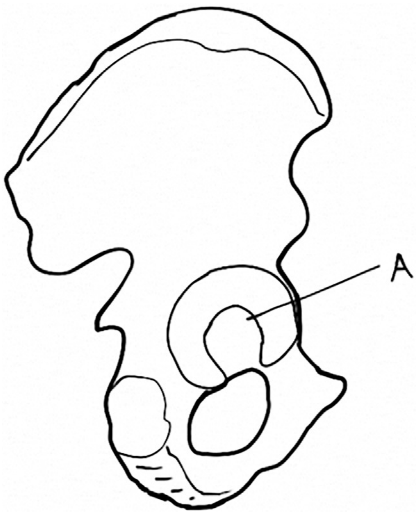


Fig. 1 Point A indicating the approximate area of the acetabular point.

accuracy of 80% was found. Interestingly, when the ischio-pubic index was used in combination with the femoral head diameter in a discriminant function (DF), the resulting classification accuracy increased to 95%. This indicates that the acetabulum, a reflection of femoral head diameter, could be a potential area of interest in regard to sexual dimorphism. Thieme and Schull's study provided a means of progress from Washburn, by utilizing DF; however, the inclusion of the acetabular point as a landmark brings into question the method's reliability and replicability.

Several other indices have been studied for use in sex estimation, including the acetabulum and pubis index (Breathnach, 1965) and the sciatic notch/acetabular ratio (Kelley, 1979). Schulter-Ellis, Schmidt, Hayek, and Craig (1983) performed a study of three indices, the acetabulum-pubis index, the acetabular diameter/pubic tubercle-acetabular rim index, and the ischium-acetabulum height/pubic symphysis-acetabular rim index. Utilizing a DF derived from the acetabulum-pubis index and the ischium-acetabulum height, followed by sorting by femoral head diameter size, yielded a correct classification accuracy of 97% (Schulter-Ellis et al., 1983). Once again, this method utilizes not just the pelvis but the femoral head diameter, indicating that the acetabulum is an area for further study in regard to sex estimation. However, forensic anthropologists usually do not have complete skeletons to work with, so methods that rely on multiple elements cannot always be utilized, leading to a decline in the usage of many of these methods. This is further reason why methods utilizing reliable, replicable measurements from the acetabulum could prove useful in sex estimation methods.

Recent advances in metric methodology

As previously mentioned, there are several areas for improvement regarding traditional metric methods, including increasing accuracy rates when used on fragmented or incomplete remains, developing reliable and replicable landmarks and measurements, and using robust and transparent statistics. Some more recent publications have attempted to remedy these areas, including the need for more clearly defined measurements. Albanese (2003) began the movement to find an alternative to the use of the acetabular point when measuring pubic length, as well as to find a method that would perform well among various populations from different time periods. A newly proposed measurement, superior pubic ramus length, was found to reduce intraobserver error and to be more reliable than the traditional measurement. Additionally, this method was found to be reliable across samples from varying populations and time periods when combining pubis length and femoral head diameter, thereby necessitating both skeletal elements.

With the recent development of new statistical analyses, and increasing technological advances, there have been several new metric pelvis sex estimation methods that have emerged. Notably, Murail, Bruzek, Houët, and Cunha (2005) created the program, known as *Diagnose Sexuelle Probabiliste* (DSP), to assess the sex of an unknown individual through metric measurements. The original DSP program (i.e., not DSP2) performs a logistic regression in a spreadsheet after 10 pelvic measurements are entered. This methodology sought to create well-defined measurements, as well as utilizing the logistic regression, which is much more robust than simple indices. While this method made much progress in terms of ushering in more rigorous methodologies, it did have its limitations. The spreadsheet provided in the original DSP program is locked and does not allow the user to see the regression formulae being used, and the regression formulae are not provided in the accompanying manuscript (Murail et al., 2005). Additionally, there are no inter- or intraobserver error rates provided, or group means for the measurements, and it is unclear how the classification accuracies provided were validated. This is a common trend among the previously discussed methods, as none use cross-validation or contain a holdout sample for internal validation to estimate the accuracy of the method when used on individuals outside the original sample. These aspects of the original DSP method do not conform to *Daubert* standards, which renders this method unpresentable in a court of law from a forensic perspective (Daubert v. Merrell Dow Pharmaceuticals, 1993). The DSP method also recommends that only individuals with high posterior probabilities (those over 0.95) should be classified based on recommendations from Franklin, Flavel, Cardini, and Marks (2013) and Kranioti and Apostol (2015). Due to this recommendation, the extremely high classification accuracies of over 98% cited only apply to the individuals deemed "classifiable" (i.e., who had a posterior probability over 0.95).

In some cases, the percentage of classifiable individuals was as low as 40%. The original DSP method, while a great improvement upon historical methods, in reality is a "black box," wherein the user inputs data and receives output, but cannot access the information and equations used to create the output. These methods would not be acceptable in a post-*Daubert* courtroom, where forensic methods must be transparent and clear with how results are achieved and how accuracy rates are calculated. Bioarchaeological contexts lack this inherent limitation; however, the black box approach can still be problematic. Nearly all these critiques were addressed with the introduction of DSP2, which was published in 2017 (see Chapter 15 of this volume for more details).

Albanese, Eklics, and Tuck (2008) progressed through building upon an earlier method (Albanese, 2003) in an effort to classify fragmented remains. To do this, the authors examined the proximal femur and created new measurements and angles that demonstrate similar discriminatory power to that of the pubis. This method utilized two logistic regression equations involving six measurements from the pelvis and proximal femur to estimate sex while also providing a probability value that the individual in question belongs to the assigned group.

Baumgarten and Ousley (2015) sought to improve upon the previously discussed original DSP method (Murail et al., 2005) while attempting to address several key issues of traditional methods. In light of the unreliability of some measurements, including the acetabular point, Baumgarten and Ousley (2015) sought to modify or create measurements with clear landmarks or measurements that are either minima or maxima. Minima and maxima measurements are known to be more replicable and easier for practitioners to use, regardless of experience, which make them ideal for use (Adams & Byrd, 2002; Jantz, Jantz, & Devlin, 2016).

Using linear stepwise DF analysis, a combination of five variables was shown to provide classification accuracies of 96% in males and 99% in females, for a pooled-sex accuracy of 97.5% in a sample of 200 individuals. The five measurements selected by the stepwise DF analysis included minimum apex to symphysion, maximum innominate length, maximum ischial length, maximum innominate breadth, and maximum pubic length. These measurements are effectively able to capture dimorphism in the innominate. Minimum apex to symphysion (Fig. 2A) captures true pelvic morphology; maximum innominate length (Fig. 2B) and maximum innominate breadth (Fig. 2C) inform on the ratio of height to width; and maximum ischial length (Fig. 2D) and maximum pubic length (Fig. 2E) illustrate the ratio of pubis to ischium.

Forty individuals were measured a second time to calculate the technical error of measurement (TEM) and coefficient of reliability for each measurement. TEM showed a less than 2 mm difference between trials and a mean of less than 3.5% for all measurements, signifying low intraobserver error rates. Additionally, for all but two of the measurements, the coefficient of reliability values was greater than or equal to 0.96, indicating an extremely high level of intraobserver consistency for those measurements.

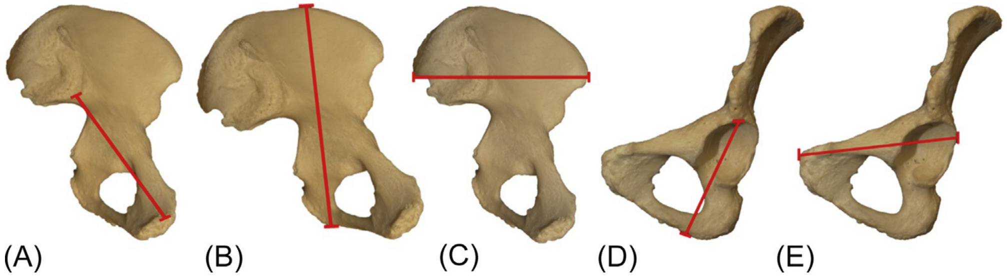


Fig. 2 (A) Minimum apex to symphysion; (B) maximum innominate length; (C) maximum innominate breadth; (D) maximum ischial length; (E) maximum pubic length.

The high levels of intraobserver agreement between rounds of measurements reveal the value of measurements with clear, unambiguous landmarks, or those that involve maxima and minima. This approach proves that current methods can be further improved upon to reduce measurement errors and increase classification accuracy.

Baumgarten, Ousley, Decker, and Shirley (2015) sought to create a method for sex estimation using these well-defined measurements that could be applied to fragmented remains, and that could be utilized on dry bone with the use of calipers as well as through virtual means. The study utilized a total of 11 measurements on the innominate, previously found to be reliable, from which six DF equations were created, based on common scenarios of fragmentation (Baumgarten & Ousley, 2015). Several of these measurements have been found to be extremely reliable and reproducible and, as such, added to the Data Collections Procedures for Forensic Skeletal Material 2.0 and FORDISC (Langley, Jantz, Ousley, Jantz, & Milner, 2016). The DF equations were created using a known sex sample of 200 innominates from the Hamann-Todd collection, housed in the Museum of Natural History in Cleveland, Ohio. Cross-validated classification accuracies for these equations were all >90%, with very low levels of sex bias (<5%).

A second sample of 150 innominates was collected from 3D volume-rendered CT scans to validate the six equations. The validated classifications displayed high levels of accuracy, with four of the equations reaching over 94% correct classification and the remaining two equations having over 83% correct classification. This validation sample also provided the opportunity for interobserver error testing, with most measurements being less than 3% error and the remaining measurements under 10% error. The validation sample provides a modern comparison to the more historic Hamann-Todd sample, demonstrating the efficacy of these equations on both historic and modern samples. These measurements have been shown to have very low intraobserver rates, with many rates under 10% (Baumgarten & Ousley, 2015). The results of this study suggest that the equations created can be used to reliably estimate the sex of unknown individuals while avoiding issues of intra- and interobserver errors, a low posterior probability (as would be unclassifiable using the original DSP). Furthermore, there is complete transparency in how the results are obtained and broad applicability to both dry bone and virtual models.

Case study

The Erie County Poorhouse skeletal collection

Background

During the 19th century, the city of Buffalo, New York, and the surrounding Erie County grew at an unprecedented rate due to an influx of immigrants seeking employment opportunities (Byrnes, 2017). This population growth occurred simultaneously with a national increase in social assistance programs, and thus, the Erie County Poorhouse (ECP) was established with the acceptance of its first residents in 1829 in the Black Rock neighborhood of Buffalo (Gerber, 1989; Muller, 2017). Quickly overflowing its original occupation limit, the poorhouse moved to the Buffalo Plains location in 1851, where it remained until its closure in 1926 (Nystrom, Sirianni, Higgins, Perrelli, & Raines, 2017). This land is currently located on the University at Buffalo (UB)'s South Campus. During its tenure, the ECP evolved into a multi-building complex containing a poorhouse, an insane asylum, a children's ward and school, a hospital with both maternity and consumptive wards, and a cemetery known as the Erie County Poorhouse Cemetery (ECPC) (Byrnes, 2017; Higgins, 1998; Higgins, Raines, & Montague, 2014).

The skeletal remains that are the focus of this case study were exhumed from the ECPC and are, therefore, known to be associated with the EPC. It is thought that the ECPC was in use for about 61 years and contains the remains of unclaimed bodies of both poorhouse inmates and hospital patients (Nystrom et al., 2017). Cemetery records indicate that 120 individuals were moved from the original Black Rock location to the Buffalo Plains location in 1852, indicating that interment at the ECPC probably began almost immediately after the relocation (Nystrom et al., 2017). Though the ECP facility was still in use up until its closure, records indicate that burials ceased in 1913 (Higgins et al., 2014). Out of the 7186 deaths that occurred between 1880 and 1913, 44% ($n = 3198$) were unclaimed and most likely buried in the ECP cemetery (Higgins et al., 2014).

In 2012, the construction of UB's South Campus prompted a large-scale excavation that unearthed the skeletal remains of 376 individuals associated with the ECP. The excavation was conducted by Archaeological Survey staff, an applied archaeology contract division from UB's Department of Anthropology (Byrnes, 2017). Though the large-scale excavation took place in 2012, there are several recorded instances of soil disturbances as a result of roadside construction. In 2009, 16 individuals were excavated due to construction at the University's childcare facilities. Further, in 1980, sanitation work was conducted in this location and is thought to have further disturbed any human remains that were buried in the area (Nystrom et al., 2017).

The 2012 excavation represents about 20% of the estimated size of the cemetery (Perrelli & Hartner, 2016). The orientation and spacing of coffins indicate a demarcation between what appears to be an older section and newer section of the cemetery. The older section contains a general observation of greater deterioration of human remains and coffins, with more soil being found inside the coffins, along with the use of older coffin hardware including nails and screws (Perrelli & Hartner, 2016). Evidence of a newer section includes preservation of clothing, better preserved remains, and wire-drawn nails, along with newspapers bearing dates between 1901 and 1903 (Perrelli & Hartner, 2016). This differential preservation pattern is important to note for the analysis of fragmented remains, as some individuals will be highly fragmented, while others are better preserved.

During the initial 2012 excavation, standard sex and age methods were used for demographic analysis of all burial locations with human remains. Of the 376 individuals excavated from the poorhouse, 58 are estimated to be under two years of age, eight are between 2 and 16 years of age, and 310 are estimated to be adults (>16 years) (Byrnes, 2015) (Table 1). Age estimates were based on the changes to the auricular surface (Buckberry & Chamberlain, 2002), pubic symphyseal morphology (Brooks & Suchey, 1990), and changes to the sternal rib ends (İşcan, Loth, & Wright, 1984a, 1984b, 1985).

Sex estimation

Sex estimates were based on morphological observations of the cranium and pelvis (Buikstra & Ubelaker, 1994) and the pubic bone (Phenice, 1969). However, due to the fragmented nature of the remains, 62.5% of the individuals could not be assigned an estimated sex. Of the 310 adults buried in the ECP cemetery, 69 were estimated to be male and 47 as female, with the remaining 194 individuals being categorized as probable male/female, ambiguous, or indeterminate (Table 1). In addition to their fragmented nature, many individuals exhibited signs of pathology throughout their skeleton,

Table 1 Age and sex distribution of ECP skeletal collection.

	Adolescent/ young adult (16–35 years)	**Middle adult (35–50 years)**	**Middle/ old adult (30+ years)**	**Old adult (50+ years)**	**Adult (20+ years)**	**Indeterminate age**	**Total**
Male	19	35	9	6	0	0	69
Prob. male	3	8	24	4	3	0	42
Ambiguous	6	3	3	1	1	0	14
Prob. female	16	5	17	2	7	0	47
Female	5	6	30	2	7	0	50
Indeterminate	9	1	23	1	37	17	88
Total	58	58	106	16	55	17	310

Modified from Nystrom, K. C., Sirianni, J. E., Higgins, R., Perrelli, D., & Raines, J. L. (2017). Structural inequaliy and postmortem examination at the Erie County Poorhouse. In: K. C. Nystrom (Ed.), *The bioarchaeology of dissection and autopsy in the United States* (pp. 279–300). New York: Springer; Byrnes, J. F. (2017). Injuries, impairment, and intersecting identities: The poor in Buffalo, NY 1851-1913. In: J. F. Byrnes, J. L. Muller (Eds.), *Bioarchaeology of impairment and disabilty: Theoretical, ethnohistorical, and methodological perspectives* (pp. 201–224). Springer.

which sometimes inhibited sex and/or age estimation, though the presence of pathology was to be expected from a population of poorhouse inmates and hospital patients (Sirianni, Higgins, & Byrnes, 2014).

To more accurately estimate sex in the most fragmented remains, the Baumgarten (2016) method was employed. This method utilizes DF equations created from modifications to existing methodologies (e.g., Albanese et al., 2008; Murail et al., 2005; Schulter-Ellis et al., 1983; Thieme & Schull, 1957; Washburn, 1949). The Baumgarten (2016) method uses 11 metric measurements, which can be taken with digital sliding calipers and then entered into one of six DF equations depending on preservation conditions (Tables 2 and 3).

The method is particularly useful for fragmented remains, which is why it was employed here. For example, if the anterior superior iliac spine (ASIS) is missing or badly eroded, then equation one is used since measurements of the ASIS are not included. All measurements corresponding to the appropriate equation are taken, recorded, and entered into the corresponding DF. The results of the equation are either a positive

Table 2 Measurements and anatomical definitions from the Baumgarten (2016) study.

Measurement	Abbreviation	Anatomical definition
Maximum innominate height	XIH	From most distal point on iliac crest and ischium
Maximum ischio-pubic ramus length	XIRL	Maximum measurement from most inferior point on pubic symphyseal face to most distal point on ischium
Maximum ischial length	XISL	Most anterior point on acetabular rim where iliac blade meets acetabulum to most medial point on ischial tuberosity
Minimum ischial length	WISL	Minimum measurement from most medial point on epiphysis of ischial tuberosity to closest point on acetabular rim
Maximum iliac breadth	XIB	Maximum measurement from posterior superior iliac spine to anterior superior iliac spine
Maximum apex to symphysion	XAS	Maximum measurement from symphysion to closest point on rum of apex of auricular surface
Minimum apex to symphysion	WAS	Minimum measurement from symphysion to most distance point on acetabular rim
Maximum pubis length	XPL	Maximum measurement from symphysion to closest point on acetabular rim
Minimum pubis length	WPL	Minimum measurement from symphysion to closest point on acetabular rim
Anterior superior iliac spine of symphysion	ASISS	Symphysion to apex of anterior superior iliac spine
Posterior superior iliac spine to symphysion	PSISS	Symphysion to apex of posterior superior iliac spine

Table 3 DF equations are used to estimate sex in a particular individual based on which features are missing.

Equation number	Missing feature(s)	Equation
1	Anterior superior iliac spine	WAS(0.464) − WIB(0.329) − WISL (0.026) + WPL(0.340) − XIH(0.418) + XIRL(0.304) − ISL(0.104) − XPL (0.104) − PSISS(0.022) + 26.803
2	Posterior superior iliac spine	WAS(0.433) − WIB(0.290) − WISL (0.024) + WPL(0.331) + XAS (0.133) − XIH(0.467) + XIRL (0.282) − ISL(0.147) − XPL(0.125) + 26.057
3	Ischium	WAS(0.310) − WIB(0.419) + WPL (0.482) − ASISS(0.021) − XIB (0.115) − XPL(0.394) − PSISS(0.006) + 22.751
4	Anterior superior iliac spine, posterior superior iliac spine, and pubic symphysis	WISL(0.027) − XIH(0.216) + XIRL (0.470) − XISL(0.207) + 18.300
5	Pubic symphysis	WIB(−0.146) + WISL(0.041) + XIB (0.220) − XIH (0.336) + XIRL (0.447) − ISL(0.213) + 19.264
6	Iliac blade	WISL(−0.141) + WPL(0.430) + XIRL (0.261) − ISL(0.394) − XPL(0.226) + 18.188

or negative number; positive numbers are classified as females, and negative numbers are classified as males.

Materials and methods

This case study examined 27 individuals from the ECPC skeletal collection, with the goal of either corroborating the existing sex estimation or suggesting a new sex estimation based on the metric analysis. This was especially promising for estimating sex of individuals who were unable to be sexed after morphological analysis. A control subsample comprised seven individuals who had undergone genetic testing in a previous study to determine the presence of a Y-chromosome (Mayberry, 2017). The left innominate was measured when possible, though 12 of the 27 individuals did not have a left innominate, in which case the right was measured. Intraobserver error was calculated on one complete innominate. Each of the 11 measurements was taken three times, averaged, and the coefficient of variation (CV) was calculated; all measurements were repeatable with a <10% CV.

Each individual innominate was evaluated for missing features, which were recorded and appropriate equations were chosen based on those recordings. Metric measurements were then taken only for the features required for each equation, and all measurements were repeated three times with a minimum of 2 weeks between each session to control for intraobserver bias. After all measurements were complete, data were entered into a Microsoft Excel file and CVs were calculated to ensure a <10% CV for each individual. The Excel file was set to calculate the DF for each individual based on an appropriate equation. DF was calculated each time it was measured, and the three resulting DF scores were averaged to estimate sex for each individual.

Results

The resulting DF numbers corroborated morphological and/or genetic sex 70.4% of the time ($n = 19$), suggested male instead of female 7.4% of the time ($n = 2$), suggested female instead of male 7.4% of the time ($n = 2$), and proposed sex for all four of the previously unknown specimens (Table 4). For the control sample of seven individuals with the presence or absence of a Y-chromosome (Mayberry, 2017), the DF, morphological, and genetic sex estimates were the same 71.43% of the time ($n = 5$). In one instance, the genetic sex was male, and the DF proposed female; and in another instance, the genetic sex was female, though the DF suggested male.

It is also noteworthy to discuss the frequencies of which equation was used, as this explains which elements of the innominate are most frequently absent in this sample (Table 5). Eq. (4), which was used when the anterior superior iliac spine, posterior superior iliac spine, and/or pubic symphysis were missing, was used most often ($n = 10$ or 37% of the time). Typically, not all of these elements were missing, but it is the only equation that accounts for a missing pubic symphysis, which was the most often missing skeletal element.

Discussion

The original study with which this method was developed produced an accuracy rate of 97.5%, which is higher than that of previous similar studies (Baumgarten, 2016). In the instances of disagreement with genetic testing and the results of this study, there are several factors that must be considered. The averaging of three DF results per individual, rather than averaging of three measurements per variable, could have led to a misclassification. Additionally, genetic testing only tested for the presence of a Y-chromosome. Due to the poor preservation of the remains, it is possible for sufficient degradation of the DNA to have occurred where the presence of the Y-chromosome was simply not detected during genetic testing (see Chapter 21 in this volume).

The original study was based on ideal conditions, with complete innominates exhibiting no obvious pathologies, which raised some question as to its application on a bioarchaeological population. It must be accounted for that this study analyzed

Table 4 Results.

Location number	Equation used	Side measured	DFA result (avg.)	DFA sex	Macroscopic evaluation sex	Genetic sex
48	4	Left	−7.009	**Male**	Unknown	
49	5	Left	−11.11	**Male**	Probable female	
51	4	Left	−16.27	**Male**	Unknown	
53	6	Right	−1.27	Male	Male	
54	2	Left	−13.86	Male	Male	
77	4	Right	−19.14	Male	Male	
79	3	Left	−0.998	Male	Male	
171	1	Left	0.469	**Female**	Male	
84	4	Right	−14.13	Male	Probable male	
220	4	Left	−23.6	Male	Male	
227	5	Left	−8.939	Male	Male	
237	1	Right	−10.12	Male	Male	
242	5	Left	−7.465	Male	Male	
268	4	Left	−18.1	Male	Male	
308	5	Right	−29.62	Male	Male	
459	3	Left	−17.25	Male	Male	
460	4	Left	−22.83	Male	Male	
464	4	Right	−22.36	Male	Male	
474	3	Right	−12.09	Male	Male	
476	3	Right	−5.252	**Male**	Female	
95-C	4	Left	−27.3	Male	Male	Male
244-C	6	Right	−5.228	Male	Male	Male
301-C	6	Right	13.9	**Female**	Male	Male
329-C	4	Right	−23.97	**Male**	Female	Female
353-C	3	Right	−19.46	Male	Male	Male
463-C5	5	Left	−7.443	Male	Male	Male
463-C2	2	Left	−4.669	Male	Male	Male

Results in bold are different than the morphological estimate or genetic sex results.

Table 5 Frequencies of DFA equations used.

Equation	Number of times used	Percentage used
1	2	7.41%
2	2	7.41%
3	5	18.52%
4	10	37%
5	5	18.52%
6	3	11.11%

individuals from a late 19th-century poorhouse and hospital population, meaning that a large majority of individuals in this population likely experienced instances of illness, malnutrition, and lack of adequate healthcare, all of which may have manifested in their skeletal remains (Byrnes, 2017).

For a poorhouse population, or another population with highly fragmented remains, this metric analysis is a fast and reliable method. On average, the measurements for each equation took about 10 min to complete and only required a set of calipers. The results from this sample suggest that this method may be more reliable than more traditional morphological analyses, as the DF results suggested a change in morphological sex estimation 17.39% of the time ($n = 4$). Furthermore, given that this metric analysis can estimate sex when morphological analysis cannot, it is of greater utility than morphological methods when used in bioarchaeological or forensic studies.

Conclusion

The innominate is commonly viewed as the best skeletal element used in the estimation of sex of an unknown individual, and morphological methods have dominated. However, metric methods can provide a more objective means of estimation. Traditional metric studies cite high accuracy rates of at least 90% range, though many of these methods use measurements based on landmarks that are difficult to find and nearly impossible to replicate, leading to high interobserver error rates. Additionally, these methods were created using complete innominates and, in several cases, involved the use of the proximal femur. This neglects the regular occurrence of fragmentation of elements and relies on both the innominate and femur being recovered, available for analysis, and well preserved.

Recently, several new metric sex estimation methods have been published that attempt to remedy the issues with traditional metric methods. These studies are promising, with many claiming accuracy rates of at least 95% in diverse samples from around the world, avoiding bias in sex classification due to ancestry. The recent publication of several new methods demonstrates progress, including the incorporation of rigorous statistical methods, incorporation of technological advances (including virtual analyses), and inclusion of methods suitable for remains that are fragmented and incomplete. Future studies will, no doubt, continue to create new metric sex estimation methods that will keep pace with the expanding technology available to anthropologists.

References

Adams, B., & Byrd, J. (2002). Interobserver variation of selected postcranial skeletal measurements. *Journal of Forensic Sciences*, *47*(6), 1193–1202.

Albanese, J. (2003). A metric method for sex determination using the hipbone and the femur. *Journal of Forensic Sciences*, *48*(2), 263–273.

Albanese, J., Eklics, G., & Tuck, A. (2008). A metric method for sex determination using the proximal femur and fragmentary hipbone. *Journal of Forensic Sciences*, *53*(6), 1283–1288.

Bass, W. M. (2005). *Human osteology: A laboratory and field manual* (5th ed.). Columbia, MO: Missouri Archaeological Society.

Baumgarten, S. E. (2016). A metric approach to sex estimation using fragmented Os Coxae. *American Journal of Physical Anthropology*, *159*, 88.

Baumgarten, S. E., & Ousley, S. D. (2015). Estimating sex from the innominate utilizing new measurements. In *Vol. 21.Proceedings of the 67th annual scientific meeting AAFS*, (p. 142).

Baumgarten, S. E., Ousley, S. D., Decker, S. J., & Shirley, N. R. (2015). A transparent method for sex estimation using refined DSP measurements of the innominate. *American Journal of Physical Anthropology*, *156*, 80.

Brauer, G. (1988). In R. Knussman (Ed.), *Osteometrie in anthropologie: Handbuch der vergleichenden biologie des Menschen. Ban 1, Wesen und methoden der anthropologie, Teil 1. Wissenschaftstheorie, Geschichte, Morphologische Methoden*. Stuttgart, Germany: Gustav Verlag.

Breathnach, A. S. (Ed.), (1965). *Frazer's anatomy of the human skeleton* (6th ed.). Boston: Little, Brown and Co.

Brooks, S., & Suchey, J. M. (1990). Skeletal age determination based on the os pubis: A comparison of Acsádi-Nemeskéri and Suchey-Brooks methods. *Journal of Human Evolution*, *5*(3), 227–238.

Buckberry, J. L., & Chamberlain, A. T. (2002). Age estimation from the auricular surface of the ilium: A revised method. *American Journal of Physical Anthropology*, *119*(3), 231–239.

Buikstra, J., & Ubelaker, D. H. (1994). *Standards for data collection from human skeletal remains*. Fayetteville: Arkansas Archeological Survey Research Series.

Byers, S. N. (2002). *Introduction to forensic anthropology: A textbook*. Boston, MA: Allyn and Bacon Publishers.

Byrnes, J. F. (2015). *'A Pauper's shame': A biocultural investigation of trauma, impairment, and disability in the Erie County poorhouse cemetery, 1851-1913*. PhD Dissertation Buffalo, NY: University at Buffalo, SUNY. 171 p.

Byrnes, J. F. (2017). Injuries, impairment, and intersecting identities: The poor in Buffalo, NY 1851-1913. In J. F. Byrnes & J. L. Muller (Eds.), *Bioarchaeology of impairment and disabilty: Theoretical, ethnohistorical, and methodological perspectives* (pp. 201–224): Springer.

Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993). In U. S. S. Court (Ed.), *United States Supreme Court*.

Franklin, D., Flavel, A., Cardini, A., & Marks, M. K. (2013). Pelvic sexual dimorphism in a Western Australian population: Integration of geometric and traditional morphometric approaches. *Proceedings of the American Academy of Forensic Sciences, LXV*, 419–420.

Gaillard, J. (1961). Valuer de l'indice ischio-pubien pour la determination sexuelle de l'os coxal. *Bulletin et Momoires de la Societe d'Anthropologie de Paris*, *2*, 92–108.

Genoves, S. (1959). L'estimation des diffe'rences sexuelles dans l'os coxal; differences metriques et differences morphologiques. *Bulletins et Memoires de la Societe d'Anthropologie de Paris*, *10*, 3–95.

Gerber, D. A. (1989). *The making of an American pluralism: Buffalo, NY, 1825–1860*. Urbana, Illinois: University of Illinois Press.

Gonzalez, P. N., Bernal, V., & Perez, S. L. (2009). Geometric morphometric approach to sex estimation of human pelvis. *Forensic Science International*, *189*(1–3), 68–74.

Higgins, R. (1998). *The biology of poverty: Epidemiological transition in Western New York*. PhD Dissertation Buffalo, NY: University at Buffalo, SUNY.

Higgins, R., Raines, J. L., & Montague, N. L. (2014). Confirming burial locations in the Erie County Poorhouse Cemetery using death certificates and mortality records from 1880-1913. *American Journal of Physical Anthropology*, *153*(S58), 141.

İşcan, M. Y., Loth, S. R., & Wright, R. K. (1984a). Age estimation from the rib by phase analysis: White males. *Journal of Forensic Sciences*, *29*(4), 1094–1104.

İşcan, M. Y., Loth, S. R., & Wright, R. K. (1984b). Metamorphosis at the sternal rib end: A new method to estimate age at death in white males. *American Journal of Physical Anthropology*, *65*(2), 147–156.

İşcan, M. Y., Loth, S. R., & Wright, R. K. (1985). Age estimation from the rib by phase analysis: White females. *Journal of Forensic Sciences*, *30*(3), 853–863.

Jantz, R. L., Jantz, L. M., & Devlin, J. (2016). Secular changes in the postcranial skeleton of American whites. *Human Biology*, *88*(1), 65–75.

Kelley, M. A. (1979). Sex determination with fragmented skeletal remains. *Journal of Forensic Sciences*, *24*(1), 154–158.

Kranioti, E. F., & Apostol, M. A. (2015). Sexual dimorphism of the tibia in contemporary Greeks, Italians, and Spanish: Forensic implications. *International Journal of Legal Medicine*, *129*, 357–363.

Langley, N. R., Jantz, L. M., Ousley, S. D., Jantz, R. L., & Milner, G. (2016). Data collection procedures for forensic skeletal material 2.0. University of Tennessee and Lincoln Memorial University.

Mayberry, M. M. (2017). *Biological anthropology in the twenty-first century: A comparative analysis of ancestry determination through osteological and genetic techniques among the human remains excavated from the Erie County Poorhouse Cemetery*. PhD Dissertation Buffalo, NY: University at Buffalo, SUNY. 126 p.

Mestekova, S., Bruzek, J., Veleminska, J., & Chaumoitre, K. (2015). A test of the DSP sexing method on CT images from a modern French sample. *Journal of Forensic Sciences*, *60*(5), 1295–1299.

Moeschler, P. (1964). *Structures morphologiques et dimorphisme sexuel: essai de différenciation métrique: application à l'os coxal*. Université de Genève. Thèse.

Muller, J. L. (2017). Rendered unfit: "Defective" children in the Erie County Poorhouse. In J. F. Byrnes & J. L. Muller (Eds.), *Bioarchaeology of impairment and disabilty: Theoretical, ethnohistorical, and methodological perspectives* (pp. 119–140): Springer.

Murail, P., Bruzek, J., Houët, F., & Cunha, E. (2005). DSP: A tool for probabilistic sex diagnosis using worldwide variability in hip-bone measurements. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, *17*(3–4), 167–176.

Nystrom, K. C., Sirianni, J. E., Higgins, R., Perrelli, D., & Raines, J. L. (2017). Structural inequaliy and postmortem examination at the Erie County Poorhouse. In K. C. Nystrom (Ed.), *The bioarchaeology of dissection and autopsy in the United States* (pp. 279–300). New York: Springer.

Perrelli, D., & Hartner, J. (2016). Erie County Poorhouse Cemetery site (UB 2756) excavation methods and results. *American Journal of Physical Anthropology*, *153*(S58), 206.

Phenice, T. W. (1969). A newly developed visual method of sexing the os pubis. *American Journal of Physical Anthropology*, *30*(2), 297–301.

Pickering, R. B., & Bachman, D. C. (1997). *The use of forensic anthropology*. Boca Raton, FL: CRC Press.

Schulter-Ellis, F. P., Schmidt, D. J., Hayek, L. A., & Craig, J. (1983). Determination of sex with a discriminant analysis of new pelvic bone measurements: Part 1. *Journal of Forensic Sciences*, *28*(1), 169–180.

Schultz, A. H. (1930). The skeleton of the trunk and limbs of higher primates. *Human Biology*, *2*, 303–438.

Sirianni, J. E., Higgins, R., & Byrnes, J. F. (2014). Paleopathology and the poor: Comparing historical records of morbidity and mortality to skeletal paleopathology in the Erie County Poorhouse. *American Journal of Physical Anthropology*, *153*(S58), 241.

Spradley, M. K., & Jantz, R. L. (2011). Sex estimation in forensic anthropology: Skull versus postcranial elements. *Journal of Forensic Sciences*, *56*, 289–296.

Thieme, F. P., & Schull, W. J. (1957). Sex determination from the skeleton. *Human Biology*, *29*(3), 242–273.

Washburn, S. L. (1949). Sex differences in the pubic bone of the Bantu and Bushman. *American Journal of Physical Anthropology*, *7*, 425–432.

CHAPTER 12

Sexual dimorphism variation in Fordisc samples

Richard L. Jantz, Stephen D. Ousley
Department of Anthropology, University of Tennessee-Knoxville, Knoxville, TN, United States

Introduction

Sexual dimorphism has a long history of study in various contexts, including in biology, where it refers to all differences between the sexes except for the sex organs. In this chapter we will concentrate on human size and shape dimorphism and its consequences in forensic applications. Height is an obvious dimorphic trait, and the dimorphism has been shown to vary among populations. Hiernaux (1964) noted the lower dimorphism of African populations compared to Europeans in height. He attributed it to sex differences in plasticity, females being better able to withstand unfavorable environmental conditions. On the other hand, Eveleth (1975) argued that height dimorphism was primarily genetic, based on the pattern of variation among world populations. Eveleth's interpretation was supported by the results of Gustafsson and Lindenfors (2004), who examined sexual dimorphism in stature from groups around the world and found that males are 7% taller and the differences are isometric, meaning that the differences are proportional to mean stature. As a practical matter in forensic anthropology, the question of why dimorphism exists is of lesser importance than understanding which aspects of skeletal morphology are sufficiently dimorphic to allow reliable sex estimation of unidentified skeletons. Also of interest in the forensic context is some idea of the patterning of dimorphism among populations.

Sex dimorphism varies among populations and, over time, within populations, and the desirability of sexing criteria derived from appropriate reference samples has been frequently pointed out. There are many analyses of various populations too numerous to review here (see Ubelaker and DeGaglia; Chapter 17 of this volume). Our focus will be on sexual dimorphism in the forensic context, specifically sexing individuals using Fordisc 3.1 (Jantz & Ousley, 2005), which uses skeletal measurements and linear discriminant function analysis (Huberty & Olejnik, 2006) to help identify skeletal remains. There are several examples calling attention to Fordisc's failure to accurately sex populations not included in its reference samples. Fordisc 2 (Ousley & Jantz, 1996) had a sex-only function based on American blacks and whites, but we recognized that Hispanic individuals

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00012-2

were turning up more and more in forensic contexts, and Hispanic males were more often misclassified as female using the sex-only function, so we did not include it in Fordisc 3. Moreover, Ramsthaler, Kreutz, and Verhoff (2007) in Germans; Manthey, Jantz, Vitale, and Cattaneo (2018) in Italians; and L'Abbe, Kenyhercz, Stull, Keough, and Nawrocki (2013) in South Africans all found that Fordisc produced biased sex classifications. These studies illustrate that even populations of the same basic ethnic extraction, i.e., European, are not equivalent, and Fordisc's North American reference sample does not generalize to Europeans or people of European extraction in other parts of the world.

Before proceeding, an often-overlooked distinction requires clarification. Variation in sex dimorphism means that the difference between male and female measurement means is larger in some groups and smaller in others. This will result in variation in classification accuracies, but not necessarily in classification bias, which is a clear difference in accuracy for males versus females. Alternatively, groups may have similar sex differences, but if the means for each sex are simply shifted, standards from one group that are applied to another group will produce classification bias. For example, Manthey et al. (2018) found that the sexual dimorphism seen in US whites and Italians was about the same, as measured by Mahalanobis D^2, yet sex classification was biased. Over 25% of Italian males classified as females using the US white function. Americans have larger glabellar projections than Italians (American means = 3.99, 2.40; Italians 3.54, 1.86 for males and females, respectively), but the sex difference (1.59 vs. 1.68, Americans and Italians, respectively) is about the same (Manthey et al., 2018).

Craniometric sexual dimorphism

Sexing Forensic Data Bank (FDB) and Thai samples using Fordisc 3.1

Fordisc contains five samples from the FDB with both sexes. These five samples along with a recent Thai sample from Khon Khan University (Yuzwa, Ousley, & Tuamsuk, 2013) were subjected to two group sex discriminant functions. Stepwise selection of variables is one method of finding the best variables that separate groups and, in this context, representing measurements that show the greatest sexual dimorphism. Forward stepwise selection first finds the best single measurement that classifies groups; then finds the next measurement in combination with the first measurement that best classifies groups; and keeps adding the best measurement one at a time until there is no improvement in some classification criterion—in the case of Fordisc, Wilks' lambda, representing group separation (Jantz & Ousley, 2005). If the assumption of multivariate normality is met, and especially after outliers are removed, the best measurements for separating the sexes should be consistently identified and allow comparison of the best discriminators among groups. The stepwise procedure in Fordisc 3.1 was used with the default Wilks' forward selection and an improvement criterion of 0.005. The results are shown in Table 1. They vary considerably in the number of measurements chosen, the rate of correct classification, and sexual dimorphism as reflected in Mahalanobis D^2. The D^2 values show that

Table 1 Sex classification for FDB groups and Thai in Fordisc 3.1 using stepwise selection.

Group	Number of measurements used	Females		Males		Total	
		N	%	*N*	%	%	D^2
Whites	9	275	91.3	489	91.6	91.49	7.17
Blacks	13	61	90.2	101	89.1	89.51	7.68
Hispanic	9	40	87.5	188	88.8	88.60	5.87
Native Am.	14	27	92.6	51	90.2	91.03	9.77
Japanese	11	122	88.5	194	84.5	86.08	4.97
Thai	19	44	65.9	66	68.2	67.30	3.08

Native Americans have the highest sexual dimorphism among these groups; Euro- and African Americans are somewhat lower and very similar to each other; Hispanics and Japanese have lower sex dimorphism; and Thais have the lowest sexual dimorphism by far, even using 19 measurements. Using somewhat different stepwise settings can produce different answers, and the Thais show consistently lowest sexual dimorphism with no combinations producing greater than 80% accuracy. These craniometric results echo findings of low sexual dimorphism in the Thai using metric and morphological cranial and postcranial data (Messer, Ousley, & Tuamsuk, 2013a; Messer, Ousley, & Tuamsuk, 2013b; Powell, Ousley, & Tuamsuk, 2013; Roth, Ousley, & Tuamsuk, 2013; Yuzwa et al., 2013).

Importantly, the statistics presented in Table 1 are not strictly comparable among groups. The number and composition of measurements differ, and the variance covariance matrix (VCVM) is specific to individual groups. It is desirable to have a common set of variables and a common VCVM to carry out other comparisons. Table 2 presents the variables chosen for each of the FDB groups. No two groups have the same subset of variables. Two variables are common to all groups, ZYB and BBH. Five others are common to five of the six groups: AUB, BNL, MDH, NLH, and OBH. These variables reflect different morphological complexes: BBH and BNL reflect anterior brain case; AUB and ZYB reflect face and vault breadth; NLH and OBH reflect aspects of face height; and MDH is unrelated to any of these complexes and is a dimorphic feature on its own. These seven variables can be regarded as a core set of dimorphic variables that will provide reasonably good sex discrimination in all groups. Correct classification on these variables ranges from 73% (Thai), 85% (Native Americans, Hispanics, and Japanese) to 90% in whites. These seven variables will be used in subsequent sections for comparative analysis.

Howells samples

One might think that Howells world database provides an opportunity to assess sex dimorphism on a worldwide basis. The problem with this idea is that most of Howells

Table 2 Measurements chosen by stepwise procedure for each group.

Variables	Whites	Blacks	Hispanic	Native Am.	Japanese	Thai
AUB	X	X	X	X		X
BBH	X	X	X	X	X	X
BNL	X	X		X	X	X
BPL		X	X			X
DKB						X
EKB			X	X		
FOB				X	X	
FOL	X			X		X
FRC				X		
GOL	X	X				X
MAB			X			X
MDH	X	X		X	X	X
NLB			X		X	X
NLH	X	X		X	X	X
OBH		X	X	X	X	X
OBB		X		X		X
OCC						X
PAC		X		X	X	X
UFBR		X			X	X
UFHT			X			
WFB	X	X		X		X
XCB					X	X
ZYB	X	X	X	X	X	X

samples were sexed visually from the cranium. It seems likely that what Howells assessed visually was, to some degree, what he subsequently measured. In any case, visual sexing will have the effect of exaggerating sex differences; males that appear female and vice versa will be assigned to the wrong sex, which will reduce the overlap in distribution. van Vark, van der Sman, and Dijkema (1989) conducted such a test on Howells samples, concluding that there was significant variation among the world's populations. The dimorphism was presumably exaggerated by Howells sex assignments.

Howells data contains only three samples for which sex is known, all from anatomical collections: Zulu, North Japan, and South Japan. The Mokapu sample was sexed with the aid of postcranial remains, so can be accepted as essentially correct. We have also included Ainu because 55 of 86 are known sex, so any bias will have to come from the presumably small number of wrongly sexed in the remaining 31, a bias we will consider acceptable for present purposes.

Howells samples, plus Euro- and African-Americans on the Fordisc Howells page, were subject to the same procedure as described above, in this case beginning with 49 variables. Variables omitted were the fractions, whose purpose is mainly to determine where the subtense should be taken, WCB, MDB, and SOS, not included in the digitizing protocol, and several others deemed minimally informative. Using Howells measurements will provide a test of whether more discriminating subsets can be extracted from the larger number of measurements. Results are shown in Table 3.

Several notable results are apparent in Table 3. Mokapu, Ainu, North Japan, and South Japan have very high D^2 values and correspondingly high correct classification rates, all over 94%. Zulu has the lowest D^2, although similar to American blacks of the 19th and 20th centuries. Both whites and blacks experience about a 2% increase in correct classification with the enlarged dataset. Perhaps the most notable result is the difference between Howells Japanese and the FDB Japanese. In Table 1, the Japanese have the lowest dimorphism and correct classification rates, while in Table 3, they have high dimorphism and correct classification exceeding 94%.

There are several possibilities to explain the disparity. Howells sample sizes are smaller than the FDB sample sizes. Stepwise procedures are known to take advantage of sampling variation in selecting variables. Related to this is the larger number of variables in Howells data that allows a broader search for discriminating variables. One variable, in particular glabellar projection (GLS), is included in the subset selected for both Howells Japanese groups as well as all but one of the other groups. This feature is well established as dimorphic. Glabella size is the best discriminator in Walker's (2008) morphological sexing system.

Table 3 Sex classification for Howells known sex or reliably sexed samples.

		Females		Males		Total	
Group	**N Vars**	***N***	**%**	***N***	**%**	**%**	**D^2**
Mokapu	9	49	98.0	50	96.0	97.0	18.67
Ainu	7	38	97.4	47	95.7	96.5	16.28
N Japan	10	32	96.9	55	96.4	96.6	17.73
S Japan	18	41	97.6	49	91.8	94.4	22.49
Zulu	18	46	89.1	55	89.1	89.1	7.68
White 20th	8	143	93.7	285	91.9	92.5	9.41
Black 20th	18	29	96.6	50	80.0	86.1	9.13
White 19th	13	72	93.1	95	91.6	92.2	11.43
Black 19th	16	75	90.7	69	88.4	89.6	9.10

Mokapu: ZYB, GLS, MDH, FRC, AUB, BPL, BNL, OCC, BRR
Ainu: ZYB, GLS, MDH, FOL, OBB, LAR, OCS
N Japan: ZYB, XCB, GLS, MDH, OBH, NPH, AUB, EKB, GOL, NLB
S Japan: ZYB, GLS, MDH, NAS, WNB, BNL, EKB, LAR, FMB, FRC, BRR, DKR, JUB, NLH, NPH, DKS, NAR, OBH, XCB, ZOR
Zulu: ZYB, GLS, AUB, GOL, FMR, FRC, ASB, MAB, JUB, DKB, NLB, WMH, NLH, OBB, FMB, PRR
White 20th: ZYB, GOL, GLS, EKB, MDH, NLH, AUB, WMH
Black 20th: MDH, DKR, OBB, ASB, AUB, XFB, ZYB, NOL, BNL, OBH, DKB, EKB, FOL, MAB, FMB, WNB, BRR, NLH
White 19th: ZYB, SSR, BAR, PAC, AUB, GLS, MDH, FMB, NLH, OBB, OCS, FRS, FMR
Black 19th: ZYB, DKR, WNB, NLH, OCS, GLS, OBH, ASB, AUB, ZOR, MDH, NPH, PAC, XCB, EKR, NLB

GLS is unlikely responsible for all of the differences between Howells and FDB Japanese. The presence of GLS does not improve American whites and blacks to any appreciable degree (GLS was not selected in 20th-century American blacks), and Zulu are not as dimorphic as Howells' other groups.

The remaining explanation concerns the VCVMs. Howells chose local populations whenever possible and, on occasion, transferred crania to his separate test sample that did not seem to fit the series. By contrast, the FDB samples are national or widely dispersed ethnic groups. American whites and blacks come from all over the United States and the original local populations from which they originate no longer structure mating to any significant degree. The Japanese sample comes from north and south Japan, and the substructure among these regions has been demonstrated (Dudzik, 2015). Furthermore, FDB groups have experienced secular change to various degrees. Since there are 50 or more years of time depth in the FDB samples, secular change will also contribute to the VCVM. We would, therefore, expect the FDB-pooled VCVM to be larger than the matrix obtained from Howells data.

Comparison of FDB and Howells samples

To explore the issues raised above regarding FDB vs. Howells data, we exported each dataset using the seven core variables described above, merged them, and imported into Fordisc as a custom dataset. Unfortunately, we could not include GLS because it is not available in the FDB data. The dataset allows Howells and FDB samples to be evaluated on a common dataset and VCVM. The individual VCVMs pass the test for homogeneity, justifying the use of pooled VCVM. It is the case, however, that pooled VCVM from Howells data is smaller than FDB VCVM (log of determinant for Howells = 14.46 vs. 16.67 for FDB; Trace for Howells = 96.85 vs. 118.84 for FDB). We will determine whether the VCVM difference affects sex dimorphism.

Table 4 shows the Mahalanobis D^2 for sex differences. The first column gives the D^2 based on the VCVM pooled over all groups; the second, the D^2 based on source-specific VCVM; Howells groups based on their pooled VCVM; and FDB groups based on their pooled VCVM. The third column gives the difference. As we would expect, Howells D^2 are increased and FDB groups are decreased when using source-specific VCVMs. However, the differences are relatively small and would make little difference in interpretation.

Returning now to variation in sex dimorphism seen in column one of Table 4, there is considerable variation, ranging from 3.94 in Zulu to 8.52 in Mokapu. The pattern reflects results found by Messer et al. (2013a, 2013b), with Pacific and East Asian groups showing higher levels of sexual dimorphism, and African groups showing lower sexual dimorphism. The low sex dimorphism in Zulu bears out the difficulty Howells (1989) says he would have had if sexing from the skulls alone. The high dimorphism in Mokapu

Table 4 Mahalanobis distances for FDB and Howells samples using pooled VCVM (first column) and source-specific VCVM (second column).

Group	Sex D^2 (VCVM all pooled)	Sex D^2 (VCVM source specific)	Diff
Native Americans (FDB)	4.81	4.67	0.14
Ainu (Howells)	7.52	8.60	−1.08
Am black (FDB)	5.84	5.74	0.10
Hispanic (FDB)	4.22	4.12	0.10
Japanese (FDB)	4.78	4.67	0.11
Mokapu (Howells)	8.52	9.51	−0.99
N Japanese (Howells)	7.09	7.94	−0.85
S Japanese (Howells)	6.06	6.83	−0.77
Am white (FDB)	5.91	5.79	0.12
Zulu (Howells)	3.94	4.09	−0.15

agrees with Howells (1989) assessment that it was an easy series to sex. It is unlikely that the high dimorphism results from skeletal sexing, since postcranial remains could be examined. Howells indicates that there were only two individuals who were problematic.

The question requiring an answer is if variation in dimorphism among the groups is significant. As Konigsberg (1991) points out, the appropriate test is the interaction term in a two-level multivariate analysis of variance. This question was addressed using SAS 9.4 Proc GLM, with the following results: Wilks' lambda $= 0.9713$, $F = 1.72$, df $= 28$, 5936.2, $p = 0.018$. We can, therefore, reject the null hypothesis that the variation in sex dimorphism among these samples is the same. We pursued the question further by comparing all pairs of groups using the test described in van Vark et al. (1989) where vectors of sex differences are compared using canonical variate (CV) scores. This results in a Chi-squared value with degrees of freedom equal to the number of variables—in this case, seven. The pairwise probabilities are presented in Table 5. As might be expected from Table 4, Mokapu and Zulu differ most from other groups. Mokapu differs from five groups at 0.05 or below (Hispanic, Japanese, American black, American white, and Zulu). Zulu also differs from five groups (Mokapu, American white, Ainu, North Japanese, and Japanese). Other notable differences occur within the FDB samples, American white differs from Hispanic, and American black from Japanese.

Visual appreciation of dimorphism variation can be seen in Fig. 1, which shows the group means on the first two CVs. CV1 reflects sex dimorphism, and some population variation, while CV2 separates populations and reflects little sexual dimorphism. The males and females of each population are connected by a line, the length of which reflects the amount of sex dimorphism, bearing in mind that the plot accounts for 72% of among-group variation, so some dimorphism could be contained in later CVs. Mokapu's large

Table 5 Probability of pairwise tests for difference in sex dimorphism. *P* values <0.05 are shown in bold.

Group	Native Am	Ainu	Black	Hispanic	Japanese	Mokapu	N Japanese	S Japanese	White
Native Am	–								
Ainu	0.322	–							
Black	**0.014**	**0.041**	–						
Hispanic	0.108	0.053	0.184	–					
Japanese	0.075	0.273	**0.023**	0.092	–				
Mokapu	0.109	0.260	**0.019**	**0.001**	**0.007**	–			
North Japanese	0.352	0.902	0.056	0.090	0.380	0.560	–		
South Japanese	0.201	0.729	0.168	0.083	0.783	0.113	0.752	–	
White	**0.014**	0.182	0.615	**0.035**	0.142	**0.040**	0.205	0.363	–
Zulu	0.064	**0.011**	0.189	0.116	**0.044**	**0.001**	**0.041**	0.178	**0.020**

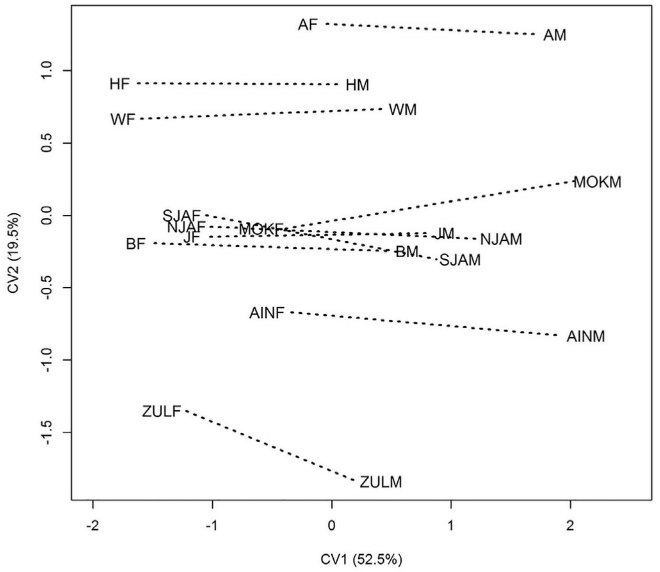


Fig. 1 CV plot of FDB and Howells groups. Lines connect sexes of the same group, which reflects the amount of sex dimorphism in each.ART: Revised figure attached

sex dimorphism and Zulu's small dimorphism is readily apparent by distances between sexes. Shifts on CV1 reflect bias. Native Americans, Mokapu, Ainu, and, to a lesser extent, Japanese are shifted toward the male end of the distribution. The canonical structure coefficients identify ZYB and AUB as the main contributors to CV1. The right-shifted groups have wider faces and vaults and are a clear indication of why population-specific criteria are normally required. CV2 is mainly concerned with population differences and expresses little sex dimorphism. It reflects mainly vault width and face height.

Dimorphism in size and shape

Differences between groups of organisms, including sexes of the same species, normally include both size and shape components. There are a variety of ways to measure size and to control it for the purpose of shape comparisons. Jungers, Falsetti, and Wall (1995) provide a review of the methods and identify Darroch and Mosimann (1985) as the best at retaining shape over different sizes. The geometric mean is isometric size. Defined in this way, size and shape are not necessarily independent, but allows examination of the role of size and shape in sex dimorphism.

It is our impression that many practitioners in forensic anthropology consider size to be the primary source of sex differences. Fordisc 3.1 has an option that permits examining the role of size and shape in sex dimorphism. Choosing the shape option computes Darroch and Mosimann's (1985) shape variables and optionally writes a dataset with shape variables and the geometric mean for each case.

Table 6 presents the mean and standard deviation for the geometric mean (size) for each group, the Mahalanobis D^2 due to size and shape. There is significant variation among groups for both sexes in size variation, but it is clear that males exhibit more size variability, as seen in their much larger *F* ratio. Comparing size and shape distances mainly supports the notion that size is the more important component of dimorphism. That is true for all groups except Native Americans and Hispanics, where the shape distance is slightly larger than the size distance.

Table 6 Summary statistics for geometric means (size), and Mahalanobis distances for size and shape alone.

	Females		Males		Mahalanobis D^2 for	
Group	**Mean**	**SD**	**Mean**	**SD**	**Size**	**Shape**
Native American	71.49	2.31	75.33	2.40	2.52	2.98
Ainu	69.89	2.10	75.00	2.30	4.48	3.79
Am black	70.19	2.15	74.97	2.38	3.92	2.67
Hispanic	69.85	2.68	73.45	3.00	2.22	2.29
Japanese	69.58	2.11	73.94	2.66	3.28	1.83
Mokapu	70.83	1.95	76.78	2.45	6.07	3.05
N Japanese	69.97	2.51	74.78	2.03	3.95	2.91
S Japanese	69.31	2.06	74.00	2.34	3.77	2.93
Am white	70.22	2.39	75.19	2.29	4.23	2.42
Zulu	68.16	2.46	71.77	2.27	2.22	1.93
F-ratio	6.70		20.78			

Table 7 Correlation between size and shape variables.

	Males	Females
AUB	−0.364	−0.344
BBH	−0.421	−0.365
BNL	−0.265	−0.244
MDH	0.534	0.442
NLH	−0.025	0.050
OBH	−0.170	−0.114
ZYB	−0.363	−0.349

As noted above, size and shape are not necessarily independent. That is the case with the present dataset. Table 7 presents the correlations between size and individual shape variables. It shows that the breadth dimensions, basion-nasion and basion-bregma, are negative. Mastoid height is positive, and the face height dimensions are low and mostly negative. This tells us that shape variation is allometric, meaning that shape varies with size. In this sample, larger crania are relatively narrower with shorter cranial bases and lower vaults than smaller crania.

Postcranial sexual dimorphism

Sexual dimorphism in human crania should be due, in part, to sexual dimorphism in body size. As mentioned, taller individuals generally show larger and longer heads, and because there are more postcranial measurements than cranial, there is a greater probability that statistical methods will better exploit any sexual dimorphism in postcranial measurements. Spradley and Jantz (2011) showed that individual postcranial metrics can be used to classify Americans, white or black, with roughly 90% accuracy, though, as with crania, one should be aware of secular changes that could affect classifications (Jantz, Meadows Jantz, & Devlin, 2016). The current study uses larger samples of American blacks and whites than were available for Spradley and Jantz (2011), and the results are quite similar to theirs. Two additional postcranial samples were analyzed for comparison, Thai (Yuzwa et al., 2013) and Hispanic. The Hispanic postcranial sample comes from positively identified border-crossing fatalities analyzed at the Pima County Office of the Medical Examiner and cemetery exhumations stored at the Universidad Nacional Autonoma de Mexico and Universidad Autonoma de Yucatan (Spradley, Anderson, & Tise, 2015).

Measurements chosen using stepwise selection of postcranial measurements in these four groups are shown in Table 8. Maximum length of the clavicle, iliac breadth, and scapula height were included for all groups. Femur head diameter, maximum humerus midshaft diameter, anterior sacrum breadth, and proximal tibia breadth were chosen in three of the four groups, largely representing robusticity, but sexual dimorphism is

Table 8 Sexual dimorphism in selected postcranial measurements. The maximum number of measurements was limited to 16 in the small Hispanic and Thai samples.

	White			Black			Hispanic			Thai		
	Number of measurements	Acc	D^2	Number of measurements	Acc	D^2	Number of measurements	Acc	D^2	Number of measurements	Acc	D^2
Major long bone lengths	6	84.8	4.90	6	88.5	5.91	6	80.0	4.18	6	70.1	1.70
Major long bone lengths (shape)	6	68.9	0.97	6	69.9	1.18	6	67.1	1.81	6	55.8	0.37
SW	10	98.4	19.81	16	97.2	44.84	12	95.2	25.45	16	94.4	18.10
SW shape	16	95.1	14.03	19	94.2	22.97	9	88.9	10.45	16	95.0	46.21
20th-century HumEpicBr[a]	1	91.4	6.68	1	91.7	6.69	1	88.3	4.37	1	90.3	5.95
19th-century HumEpicBr[b]	1	93.8	8.42	1	87.5	5.47	–	–	–	–	–	–
femhdd	1	90.1	6.25	1	87.9	6.14	1	88.4	5.00	1	82.6	4.06
tibpeb	1	91.6	7.54	1	89.7	7.86	1	89.9	3.71	1	79.4	2.90
humhdd	1	91.0	6.73	1	90.4	6.49	1	81.8	3.15	1	89.7	5.54
FEMEBR	1	90.4	5.77	1	89.2	6.22	1	79.4	2.99	1	88.3	5.98
SCAPHT	1	90.8	6.45	1	89.9	6.32	1	90.7	4.29	1	87.3	4.03
CLAXLN	1	84.1	4.17	1	87.8	4.54	1	85.2	4.13	1	85.0	3.36
ILIABR	1	63.3	0.41	1	69.1	0.88	1	67.8	0.62	1	59.0	0.25

[a] Means for epicondylar breadth (HumEpicBr) of the humerus: 20th-century black females and males: 55.3, 64.5; white: 55.1, 64.7.

[b] Means for epicondylar breadth of the humerus: 19th-century black females and males: 58.7, 66.6; white: 56.0, 64.9.

expressed differently in each group: 15 of the 30 postcranial measurements selected were used in only one group, far more than the four single craniometrics chosen.

Classification accuracies for all groups using postcranial measurements are shown in Table 9. On the whole, postcranial accuracies are higher than cranial accuracies and are more consistent among groups, with accuracies in the Thai much closer to other groups. Using six long bone measurements results in accuracies close to those for stepwise-selected craniometrics in Table 1. These six measurements represent important size and shape variation, because when the shape variables are used, the accuracy is much lower. Using stepwise selection, roughly the same number of postcranial as cranial measurements were chosen in each group, but accuracies were higher for postcranial

Table 9 Stepwise selection of postcranial measurements in four groups. The maximum number of measurements was limited to 16 in the small Hispanic and Thai samples.

Measurement	White	Black	Hispanic	Thai
CALCXL			X	
CLAXLN	X	X	X	X
FEMBLN		X		X
FEMCIR		X		
FEMEBR			X	X
FEMHDD	X		X	X
FEMMAP				X
FEMMTV		X		
FEMXLN			X	
HUMEBR	X		X	
HUMHDD		X	X	
HUMMWD	X			
HUMMXD	X	X		X
HUMXLN				X
ILIABR	X	X	X	X
INNOHT			X	
RADTVD				X
RADXLN		X		
SACABR	X	X	X	
SACAHT		X		X
SACS1BR				X
SCAPBR		X		
SCAPHT	X	X	X	X
TIBDEB		X		
TIBNFT	X			X
TIBNFX				X
TIBPEB	X	X	X	
ULNCIR		X		
ULNXLN		X		X
Unique to Grp	1	6	3	5

combinations (with 96% average accuracy) than for cranial (with 84% average accuracy). Interestingly, stepwise-selected shape variables are nearly as accurate as stepwise-selected measurements, though many outliers were discovered and removed during shape variable selection. The stepwise-selected measurements were quite similar to those selected by Tise, Spradley, and Anderson (2013) and Spradley et al. (2015).

In terms of single measurements, postcranial metrics also show greater sexual dimorphism. In fact, one postcranial measurement, epicondylar breadth of the humerus, separates 20th-century American white or black males and females over 90% correctly, and works for 19th-century Americans as well with little or no adjustment. Epicondylar breadth shows slight changes in group means and is one measurement that has been little affected by secular changes in the 19th to 20th centuries, unlike long bone lengths (Jantz et al., 2016). Epicondylar breadth and most other single postcranial measurements classified the Hispanic and Thai samples with accuracies near 90%. In contrast, the best single craniometric accuracies for American whites and blacks were seen using ZYB (82.4% and 83.1%, respectively) and GOL (78.4% and 76.4%), which pale in comparison to all postcranial accuracies except for iliac breadth in Table 8, and are lower than nine postcranial measurements in Spradley and Jantz (2011). Thus, in every group that we examined using several approaches, sexual dimorphism is greater in postcranial measurements than in cranial measurements.

Conclusions

This chapter has provided insight into sexual dimorphism variation in Fordisc's samples and its consequences for classification. Following are the major findings with a consideration of their implications:

Howells vs. FDB

Except for the Zulu, Howells samples are more dimorphic than FDB samples. That is because Howells data contains a broader range of population samples. Mokapu is the most dimorphic group; and Zulu, the least. FDB groups, by contrast, are limited to those that are likely to be found in the contemporary US population. It may also be important that Howells groups are all 19th century or earlier. Howells' Mokapu sample is precontact, and we do not know if contemporary Hawaiians would be similarly dimorphic. Another interesting Howells-FDB comparison is that the sexual dimorphism of American blacks is more similar to American whites than to the African Zulu. Also, Howells' Japanese are more dimorphic than the FDB sample, but not significantly so (compare Tables 3 and 4).

Use of stepwise in variable selection

We have used stepwise to choose variables that are most dimorphic. Stepwise is well known to take advantage of sampling variation and may produce variable subsets that are not replicable (Thompson, 1995). That may be, in part, responsible for the large number of variables chosen for most of Howells groups; samples sizes are generally smaller than in the FDB. We were able to compare the forensic and Howells groups directly by choosing the most replicable measurements in each. Our goal was to identify variables that could be used as a common set for comparative purposes. But it is important to emphasize that stepwise selection should be used with caution, particularly with small samples. As a practical matter, it may not be necessary to use stepwise for FDB samples in most cases because overfitting is not the issue we once maintained (see Fordisc 3 help file).

Size and shape

The relationship between cranial size and shape is more complicated than we might have thought at the outset. Primarily it suggests areas that could be researched more thoroughly. The relatively high correlation between size and shape variables suggests the allometric *nature* of cranial change. It has already been shown that secular change in American white crania is mainly shape (Jantz & Meadows Jantz, 2016), but its relationship to size bears further investigation. Also worth further investigation is the relationship between size and sex dimorphism, as stated in Rensch's rule (Smith & Cheverud, 2002). Messer et al. (2013a, 2013b) found support for Rensch's rule in Howells full data set, but not when using only known sex samples. The present data suggest that the matter might be more profitably investigated by limiting it specifically to size, rather than the overall distance between sexes. Male size and size dimorphism (Table 6) are highly correlated (0.786, $P = .007$), but female size and dimorphism have a lower correlation (0.354), which is not significant. The present samples are not adequate to support generalizations, but suggest that additional investigations would be informative.

Size and shape in postcranial data are important for most accurate classifications, though within-group sex classification accuracy using postcranial shape variables is nearly as high as using size and shape in the original measurements. It is, therefore, important to identify group membership before sex because of the differing patterns of size and shape in different groups. The Thai, for example, are known to show shape differences compared to other groups in morphological innominate traits (Klales, Ousley, & Vollner, 2012; Powell et al., 2013), in cranial robusticity traits (Roth et al., 2013; Walker, 2008), and in size and shape using craniometrics (Yuzwa et al., 2013). For classifications involving more groups, cranial size and shape are important; in fact, a four-way classification of black and white females and males is 90.4% accurate using stepwise selection of 15 measurements. This accuracy is essentially the same as sex classification within those groups.

Postcranial vs. cranial data

We have demonstrated that postcranial measurements classify sex better than craniometrics, as has been demonstrated in the same or similar groups before (Işcan, Loth, King, Shihai, & Yoshino, 1998; King, 1997; Spradley et al., 2015; Spradley & Jantz, 2011; Tise et al., 2013; Yuzwa et al., 2013). When analyzing remains from a group with low cranial sexual dimorphism, such as the Thai, using more accurate postcranial measurements is especially important. Standards based on the familiar Terry, Todd, and Bass skeletal collections can perform poorly when applied to other groups (Roth et al., 2013; Tallman, 2019; Tise et al., 2013). The most accurate analysis of human remains depends on having the best dataset from the most appropriate humans.

Acknowledgments

The authors thank Kate Spradley at Texas State University, San Marcos, for sharing additional data used for the Hispanic sample; and Panya Tuamsuk, Khon Khan University, for allowing us to collect Thai data.

References

Darroch, J. N., & Mosimann, J. E. (1985). Canonical and principal components of shape. *Biometrika*, *72*, 241–252.

Dudzik, B. (2015). *Investigating cranial variation in Japanese populations using geometric morphometrics*. PhD Dissertation University of Tennessee. 223 pp.

Eveleth, P. B. (1975). Differences between ethnic groups in sex dimorphism of adult height. *Annals of Human Biology*, *2*(1), 35–39.

Gustafsson, A., & Lindenfors, P. (2004). Human size evolution: No evolutionary allometric relationship between male and female stature. *Journal of Human Evolution*, *47*, 253–266.

Hiernaux, J. (1964). Weight/height relationship during growth in Africans and Europeans. *Human Biology*, *36*, 273–293.

Howells, W. W. (1989). *Skull shapes and the map*. Cambridge, MA: Peabody Museum of Archaeology and Ethnology, Harvard University.

Huberty, C. J., & Olejnik, S. (2006). *Applied manova and discriminant analysis*. New York: Wiley and Sons.

Işcan, M. Y., Loth, S. R., King, C. A., Shihai, D., & Yoshino, M. (1998). Sexual dimorphism of the humerus: A comparative analysis of Chinese, Japanese and Thais. *Forensic Science International*, *98*, 17–29.

Jantz, R. L., & Meadows Jantz, L. (2016). The remarkable change in Euro-American cranial shape and size. *Human Biology*, *88*(1), 56–64.

Jantz, R. L., Meadows Jantz, L., & Devlin, J. L. (2016). Secular changes in the postcranial skeleton of American whites. *Human Biology*, *88*(1), 65–75.

Jantz, R. L., & Ousley, S. D. (2005). *FORDISC 3: computerized forensic discriminant functions. Version 3.1*. Knoxville: The University of Tennessee.

Jungers, W. L., Falsetti, A. B., & Wall, C. E. (1995). Shape, relative size, and size adjustments in morphometrics. *Yearbook of Physical Anthropology*, *38*, 137–161.

King, C. A. (1997). *Osteometric assessment of 20th century skeletons from Thailand and Hong Kong*. Dissertation-Florida Atlantic University.

Klales, A. R., Ousley, S. D., & Vollner, J. M. (2012). A new method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *American Journal of Physical Anthropology*, *149*, 104–114.

Konigsberg, L. K. (1991). An historical note on the t-test for differences in sexual dimorphism between populations. *American Journal of Physical Anthropology*, *84*, 93–96.

L'Abbe, E. N., Kenyhercz, M., Stull, K. E., Keough, N., & Nawrocki, S. (2013). Application of Fordisc 3.0 to explore differences among Crania of North American and South African blacks and whites. *Journal of Forensic Sciences*, *58*(6), 1579–1583.

Manthey, L., Jantz, R. L., Vitale, A., & Cattaneo, C. (2018). Population specific data improves Fordisc's performance in Italians. *Forensic Science International*, *292*. 263.e261–263.e267.

Messer, D. L., Ousley, S. D., & Tuamsuk, P. (2013a). Human craniometric sexual dimorphism and Rensch's rule. In *Poster presented at the 65th annual meeting of the American Academy of Forensic Sciences, Washington, DC, February 18–23*.

Messer, D. L., Ousley, S. D., & Tuamsuk, P. (2013b). Human craniometric sexual dimorphism and Rensch's rule. In *Poster presented at the 82nd annual meeting of the American Association of Physical Anthropologists, Knoxville, TN, April 10-13*.

Ousley, S. D., & Jantz, R. L. (1996). *FORDISC 2.0: Personal computer forensic discriminant functions*. Knoxville: The University of Tennessee.

Powell, E., Ousley, S. D., & Tuamsuk, P. (2013). Sex estimation in a modern Thai sample using non-metric traits of the innominate. In *Poster presented at the 65th annual meeting of the American Academy of Forensic Sciences, Washington, DC, February 18–23*.

Ramsthaler, F., Kreutz, K., & Verhoff, M. A. (2007). Accuracy of metric sex analysis of skeletal remains using Fordisc (R) based on a recent skull collection. *International Journal of Legal Medicine*, *121*(6), 477–482.

Roth, M., Ousley, S. D., & Tuamsuk, P. (2013). Sex estimation using non-metric traits in Thai Crania with the Walker (2008) method. In *Poster presented at the 65th annual meeting of the American Academy of Forensic Sciences, Washington, DC, February 18–23*.

Smith, R. J., & Cheverud, J. M. (2002). Scaling of sexual dimorphism in body mass: A phylogenetic analysis of Rensch's rule in primates. *International Journal of Primatology*, *23*(5), 1095–1135.

Spradley, M. K., Anderson, B. E., & Tise, M. L. (2015). Postcranial sex estimation criteria for Mexican Hispanics. *Journal of Forensic Sciences*, *60*, S27–S31.

Spradley, M. K., & Jantz, R. L. (2011). Sex estimation in forensic anthropology: Skull versus postcranial elements. *Journal of Forensic Sciences*, *56*, 289–296.

Tallman, S. D. (2019). Cranial nonmetric sexual dimorphism and sex estimation in East and Southeast Asian individuals. *Forensic Anthropology*, *2*(4), 1–8.

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: a guidelines editorial. *Educational and Psychological Measurement*, *55*, 525–534.

Tise, M. L., Spradley, M. K., & Anderson, B. E. (2013). Postcranial sex estimation of individuals considered Hispanic. *Journal of Forensic Sciences*, *58*, S9–S14.

van Vark, G. N., van der Sman, P. G. M., & Dijkema, J. (1989). Some multivariate tests for differences in sexual dimorphism between human populations. *Annals of Human Biology*, *16*, 301–310.

Walker, P. L. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, *136*, 39–50.

Yuzwa, T., Ousley, S. D., & Tuamsuk, P. (2013). Sexual dimorphism in the Thai postcranial skeleton. In *Poster presented at the 65th annual meeting of the American Academy of Forensic Sciences, Washington, DC, February 18–23*.

CHAPTER 13

Statistical approaches to sex estimation

Alexandra R. Klales[a], **Stephen D. Ousley**[b], **Nicholas V. Passalacqua**[c]
[a]Forensic Anthropology Program, Department of Sociology and Anthropology, Washburn University, Topeka, KS, United States
[b]Department of Anthropology, University of Tennessee-Knoxville, Knoxville, TN, United States
[c]Anthropology and Sociology Department, Forensic Anthropology Program, Western Carolina University, Cullowhee, BC, United States

Introduction

The basis for sex estimation is the usage of data from known individuals to recognize and quantify sexually dimorphic patterns in variables between males and females. Those patterns are then applied to an unknown individual for classification in which a sex estimate is necessitated. Because sex estimation is typically a binary outcome, and considered a discrete variable (e.g., nonoverlapping categories), the family of statistics that the analyses fall within are classification techniques. Various classification approaches have been used to define and model these patterns and to generate classification algorithms for unknown individuals, with different statistical measures of certainty. These statistical methods have included classic parametric classification techniques: different forms of discriminant function analysis (DFA) such as linear, quadratic, and flexible DFA; logistic regression (LR); Naïve Bayesian (NB); and simple neural networks. Nonparametric techniques include kernel probability density (KPD), k-nearest neighbor (*k*NN), and more flexible clustering techniques such as latent profile analysis (LPA) for continuous variables, and latent class analysis (LCA) for categorical data. DFA and LR have been the time-honored approaches in biological anthropology; however, more recently, we have seen a surge in geometric morphometric analyses (GMA) and machine learning (ML) techniques. These newer ML approaches include decision trees (DT)/random forest models (RFM), artificial neural networks (ANN), and support vector machines (SVM). Recently, the optimized summed scored attributes (OSSA) method (which is essentially a sectioning point method) has also been applied to sex estimation (Mizell, Long, & Klales, 2019; Tallman & Go, 2018) with limited success.

Despite a multitude of published statistical approaches to sex estimation, many of these methods either lack associated probabilities of sex classification (e.g., trait presence/absence in Bruzek, 2002) and/or, despite high classification probabilities, are not translatable into a method that others can practically apply (e.g., GMA in Bytheway & Ross, 2010). The aims of this chapter are to provide a brief summary of the evolving approaches to sex estimation from human skeletal remains using both morphological and metric data, as well as to discuss and potentially resolve the *Measurement-Statistics*

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00013-4

Controversy. Although not exhaustive, Krishan et al. (2016) provide a more detailed description of the different statistical approaches and skeletal regions utilized for sex estimation and also include specific research papers for each, which is beyond the scope of this work.

A review of statistical approaches to sex estimation

Morphological

As far back as the 19th century, practitioners would estimate sex via bone count or, more commonly, based on their experience and exposure to sexual dimorphism using the presence or absence of specific traits or trait combinations. Today, sex estimation based on morphological traits typically relies on a presence/absence, majority rule, or the post hoc selection of traits that correspond to the sex interpretation of the investigator. This approach has been rightly criticized for its lack of statistical support and scientific rigor (cf. Hefner, 2009; Klales, Ousley, & Vollner, 2012). Specifically, this approach is considered invalid because it is subjective, typological, and fails to provide probabilities for the sex estimate (i.e., error rates). Furthermore, very few blind tests have been conducted in an attempt to verify the claimed accuracy rates of specific traits or trait combinations, perhaps with the exception of Phenice's (1969) traits. In some cases, specific features such as mandibular ramus flexure (Loth & Henneberg, 1996) continue to be utilized for sex estimation despite being independently discredited by multiple studies (e.g., Donnelly, Hens, Rogers, & Schneider, 1998; Galdames et al., 2008). Both Rogers and Saunders (1994) and Williams and Rogers (2006) attempted to rank and standardize morphological traits of the skull and pelvis for sex estimation; however, practitioner preferences indicate that the traits ranked high for precision and accuracy by these two studies are not necessarily the traits selected to make a final estimate of sex (Klales, 2013). To rectify some of the shortcomings of post hoc and inconsistent trait selection, popular traits were modified to include more objective scoring of those traits through the use of ordinal scales as well as integration of statistical analyses for classification (cf. Klales et al., 2012; Walker, 2008). Within these ordinal scales, the variation of a trait is graded from most gracile (least robust) to most robust (least gracile) expression. The order implies increased robusticity; however, the scale difference between each score is inherently inconsistent and uneven due to the nature of the level of measurement (the opposite of which is true for interval scales). As the name implies, the "order" of the parameters is most important with ordinal data/scales, rather than the degree of difference between the gradations.

Most often, LR is the preferred statistical approach of analyzing ordinally scored morphological traits, but other classification approaches (e.g., DFA) have also been frequently applied. LR calculates the probability that an individual belongs to a specific group (e.g., females or males), while DFA determines which group the unknown individual most likely belongs to, based on the overall similarity after maximizing group separation.

Both LR and DFA are popular approaches for estimating sex; however, DFA is typically restricted to metric analyses, while LR is preferred for morphological analyses. This is due to the nature of the data (continuous vs. discrete) and the associated assumptions of each statistical approach.

DFA is a statistical method that separates groups by maximizing among-group variation and accommodating correlations among measurements through a combination of weights for each variable. DFA uses those weights to classify an unknown individual into one of the known reference groups based on the Mahalanobis distance of the unknown to the known reference group's centroid (mean) (Tabachnick & Fidell, 2012). Assumptions for DFA include normal distribution and homogeneity of variances-covariance matrices. This approach is sensitive to outliers, and in order to avoid overfitting the model, the sample size must be large enough to exceed the number of predictor variables included in the function, which should be at least three to four times the number of independent variables (Tabachnick & Fidell, 2012). Linear DFA (LDFA) utilizes a linear combination of predictor variables, while quadratic DFA utilizes a nonlinear combination and does not assume equal variance between variables (Fig. 1). Each of the discriminant function techniques has assumptions regarding the data, which are not typically met with the ordinal data common in popular sex estimation methods.

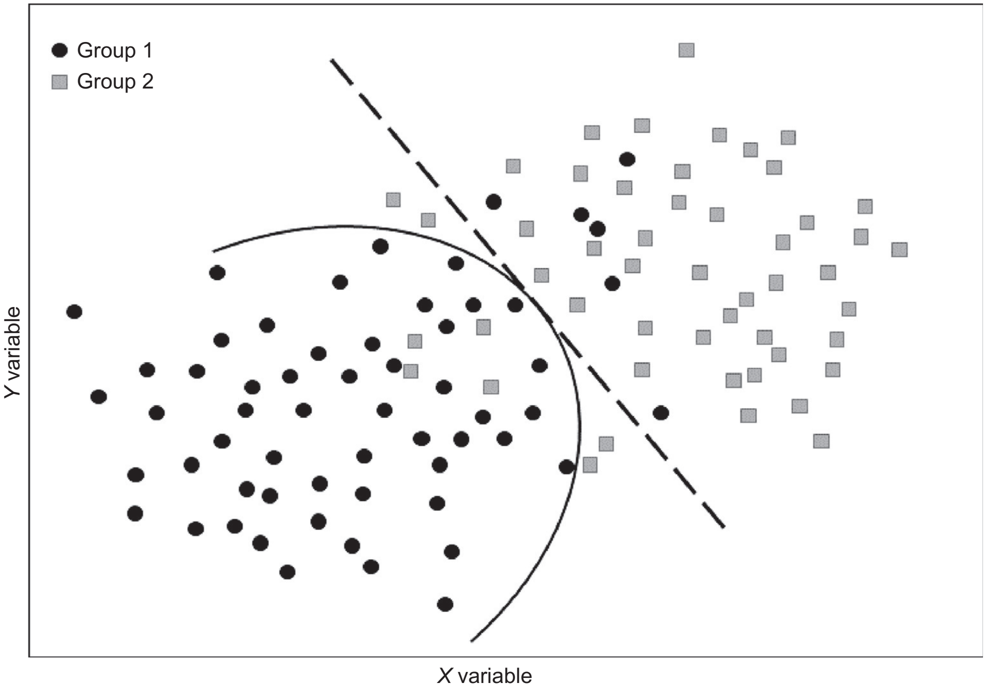


Fig. 1 LDFA (dashed black line) versus QDFA (solid black line).

As a semiparametric statistic, LR has fewer assumptions than DFA, as it does not require normally distributed data, linearly related predictor variables, or homoscedasticity (i.e., homogeneity of variance), which can make it more widely applicable to sex estimation. However, LR does require generally large sample sizes. In LR, the probability of sex membership is based on a linear combination of predictor variables (i.e., trait scores) using log transformation of predicted odds ratio (Tabachnick & Fidell, 2012). Rather than classifying the unknown into a group as in DFA, LR directly calculates posterior probabilities for each estimate (in this case, female or male) based on the relationship between the predictor variables and a binary dependent variable. In cases where the dependent variable is not binary, ordinal LR or multinomial LR can be applied. LR is more flexible than DFA and works well with both categorical data (e.g., sex or age) and continuous data (e.g., measurements).

*k*NN and KPD analyses are termed nonparametric methods because they do not require normality. These methods do not classify individuals based on similarities to reference *groups*; instead these statistical methods classify individuals based on similarities to reference *individuals* and are, therefore, considered unsupervised (i.e., classification is based on patterns found in the data rather than using predetermined labels). Both approaches are based on recognizing patterns within the data and are frequently applied to metric data. In *k*NN analysis, *k* represents the most similar individuals in the reference sample, and the classification of an unknown individual is based on group identities of those similar individuals in the reference sample. The nearness of the neighbor is based on Euclidean distance, which, unlike the Mahalanobis' distance used in DFA, does not account for variable correlation. Correct classification rates can vary with the selection of the number of neighbors (*k*) (Fig. 2) and because correlated variables become weighted more heavily in the calculation. *k*NN is considered a "lazy learning algorithm" because it does not base predictions from a training data model beforehand, but rather uses the training data input *after* a query to produce a prediction (i.e., it does not use the training data to generalize overall about that data and, therefore, a classification model is not generated from the training data until a query is made) (Aha, 1997). Each classification instance produces a new prediction model from the training data based on the number of neighbors and is hence termed lazy because the processing of training data for pattern recognition is delayed until a classification of an unknown is necessitated.

KPD analysis also avoids making assumptions about data distribution in groups. Using the training or reference data set, KPD smooths each of the individual values into a smoothed probability distribution. In areas with many individuals or observations, the density of this smoothed bump is high, but conversely so, areas with few observations where the density of smoothed bump will be low (Chen, 2017) (Fig. 3). The probability density of an individual is calculated for each group based on summing the kernels, or miniature smoothed probability densities, of each individual in the reference sample. As with *k*NN, the classification depends on the parameters selected for the analysis.

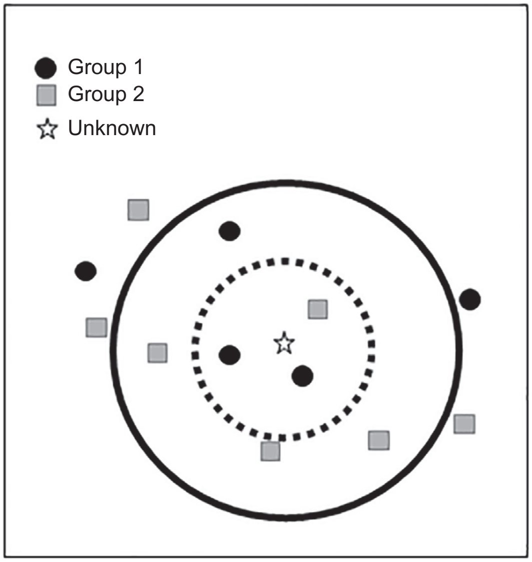


Fig. 2 *k*NN analyses. If $k=3$, the unknown individual classifies as group 1. If $k=7$, the unknown individual classifies as group 2.

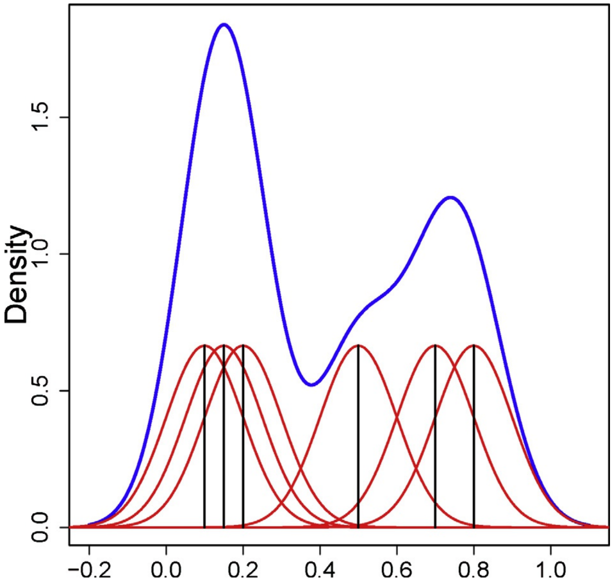


Fig. 3 Example of how individual values are smoothed using KPD. *(Figure obtained from Chen, Y. C. (2017). A tutorial on kernel density estimation and recent advances.* Biostatistics & Epidemiology, 1, *161–187, copyright © 2016, Taylor & Francis and International Biometric Society—Chinese Region, reprinted by permission of Taylor & Francis Ltd, http://www.tandfonline.com on behalf of 2016 Taylor & Francis and International Biometric Society—Chinese Region.)*

Group probability densities and classification accuracy vary with the choice of kernel type and radius, which ultimately affect smoothing.

Finally, OSSA is a heuristic method based on trait frequencies in two groups. The expressions of ordinal discrete traits are dichotomized based on frequencies found within the reference/training samples, then summed into a composite score. A sectioning point of the composite score for classification is determined based on the summed score accuracy rates in the reference sample. This simple method was originally developed for six cranial traits for ancestry estimation (Hefner & Ousley, 2014), but has since been applied to sex estimation using Walker's (2008) ordinally scored traits of the skull (Tallman & Go, 2018). Tests of the OSSA method for sex using the Walker (2008) traits indicated the method is not valid (Mizell et al., 2019).

According to Stevens's (1946) permissible statistics, LR is the most appropriate statistical method for classifying individuals based on morphological traits, but when compared to other classification methods, LR is not often the best *performing* model. Konigsberg and Hens (1998) compared the accuracy of LR (treating the data as categorical) to multivariate cumulative probit models for ordinal cranial traits. The multivariate cumulative probit models provided higher accuracy and less sex bias. Using a simulation study, Pohar, Blas, and Turk (2004) found LR to be an acceptable replacement for DFA in cases where the assumptions of DFA were violated, but the results of the study found LR to perform either about the same or just below the results obtained using DFA. Walker (2008) utilized multiple statistical approaches for classification of ordinal traits of human skulls. Accuracy was the highest using *k*NN (92.1% vs. 87.9% using LR). Klales et al. (2012) also tested multiple classification methods (linear and quadratic DFA, LR, *k*NN, and KPD) for their revision of Phenice's (1969) traits and found that some of these methods outperformed LR; however, reviewers of the manuscript recommended removal of these methods due to the violation of statistical assumptions; these were, therefore, omitted from the final publication (Vollner, Klales, & Ousley, 2009). When assumptions of a method are violated, there is a risk of presenting invalid results, which is why people choose the statistics where the assumptions are *not* violated. At present, LR remains the preferred statistical method for morphological traits, despite higher classification accuracies of other statistical methods. Also, as Konigsberg and Frankenberg (2019, p. 385) recently pointed out, statistically it is more appropriate and accurate to use sex classification traits as dependent variables and sex as the binary independent variable using ordered probit or logit regression in the spirit of "transition analysis" (Boldsen, Milner, & Hylleberg, 2002; Milner & Boldsen, 2012), rather than using logistic regression with skeletal traits as independent variables.

Metric

Metric analyses of sex initially focused on linear measurements, indices, or angle measurements that primarily captured the size differences between females and males (e.g., Stewart, 1942; Washburn, 1948). As statistical approaches and computing capabilities have evolved, there has been a move away from these simple approaches toward multivariate data using more complex statistical methods that provided sectioning points for classification. While this transition began in the 1960s with Giles and Elliot (e.g., Giles & Elliot, 1963), biological anthropologists have long since relied on "substituting individual case measurements into the corresponding DFA or regression equations, or by consulting tables of cut-off values or confidence intervals for observed traits, rather than by actually performing any statistical analysis" (Dirkmaat & Cabo, 2012, p. 12). Dirkmaat and Cabo (2012, p. 12) outline the problems with this approach, specifically the lack of associated probabilities for sex estimation: "in other words, the forensic anthropologist could not distinguish a case with a 51%–49% relative probability from a 99% to 1% case." The advent of programs such as FORDISC 3.0 (FD3) (Jantz & Ousley, 2005) and 3D-ID (Slice & Ross, 2009) alleviate this probability issue with existing metric methods. FD3 has remained the most popular method of metric sex estimation since its release (Klales, 2013) (see Chapter 12 of this volume).

Since the 2000s, sex-related differences in skeletal form (size and shape) have been explored through the use of two- or three-dimensional Cartesian coordinate data and GMA. GMA use *x*, *y*, and *z* coordinates of landmarks, semilandmarks, curves, and outlines. The coordinates are then scaled using Generalized Procrustes superimposition to uniformly translate, rotate, and scale the points (see Adams, Rohlf, & Slice, 2004 for more information) and can be used to explore shape differences, independent of size. Features of the skull (Franklin, Freedman, Milne, & Oxnard, 2006; Perlaza, 2014; Kimmerle, Jantz, Konigsberg, & Baraybar, 2008; Kustár et al., 2013; Murphy & Garvin, 2018) and pelvis (Anastasiou & Chamberlain, 2012; Bytheway & Ross, 2010) have been explored most extensively, while more recently, other postcranial bones have been evaluated for their utility in sex estimation (e.g., Brzobohatá, Krajíček, Velemínský, Poláček, & Velemínská, 2014). These methods provide a more thorough understanding of shape differences between males and females but currently have limited applicability for practical applications in sex estimation. For example, Bytheway and Ross (2010) achieved a classification accuracy of nearly 100% with GMA of the innominate using the Terry collection; however, a practitioner would be unable to utilize the method because they would need to digitize the landmarks in their unknown case and then compare it to the samples utilized within the article, which are not currently available and are likely impacted by secular change. Furthermore, GMA approaches to sex estimation may be hampered unless centroid size is also used in classification, because most of the variation between sexes appears to be due to size (Ousley & Kenyhercz, 2013). Current options for shape analyses and sex estimation include FD3 and 3D-ID (Slice & Ross, 2009). In FORDISC, shape transformations can be performed using "the Darroch and Mosimann (1985) method of conversion into shape variables whereby all original measurements are scaled by their geometric mean" (Jantz & Ousley, 2005). 3D-ID

(Slice & Ross, 2009; Ross, Slice, & Williams, 2010) allows practitioners to utilize GMA for sex estimation of the crania; however, similar programs for other skeletal elements have not yet been developed.

Machine learning

The most recent trend in sex estimation from human skeletal remains is not in the data being collected but how the data are being analyzed. Specifically, data are being explored with a variety of ML approaches with roots almost 100 years old that have been extensively developed in the last three decades, thanks to faster and more programmable electronic computers (Berry, Johnston, & Mielke, 2014), but that have only been applied within biological anthropology within the last decade. Two main types of ML exist: unsupervised (clustering and association) and supervised (including regression and classification) methods. In unsupervised learning, the machine attempts to find inherent patterns within the input data without output variables. Sex estimation relies on supervised ML. Contrary to the "lazy learning" models described above, supervised ML approaches are considered "eager learning" models in that they compile the inputs (training data) that are often resampled or transformed, and develop a model based on the rules, decisions, or networks (an algorithm) that classify best (Aha, 1997). The best performing model is determined through estimating predictive accuracy, which is based on repeated (often bootstrapped) subsampling of training data, which is tested against individuals not in the training data (test data, or holdout sample). Thousands of resampled data sets are used to estimate consensus classifications with well-established accuracy estimates. Supervised ML relies far more on prediction algorithms using patterns in the data and "rely on empirical capabilities," rather than on statistical inference and creating mathematical models based on the reference data (Bzdok, Altman, & Krzywinski, 2018, p. 234). ML stresses on classification accuracy and predictive analytics over tests of significance and descriptive data analysis. Benefits of some ML approaches for sex estimation are numerous and include: no requirement for normally distributed data, the ability to generate values for missing data, fewer statistical assumptions, and the ability to effectively use different data types (binary, nominal, ordinal, and continuous) as well as data types that do not fit Stevens' measurement scales. ML can largely make the *Measurement-Statistics Controversy* irrelevant (discussed in more detail below).

Thus far, popular supervised ML approaches applied in sex estimation include decision ST/RFM, ANN, LPA, and SVM. ML statistical approaches have also shown great promise for morphological traits (binary, discrete, ordinal data) and combined morphological/metric ancestry estimation (Hefner, Spradley, & Anderson, 2014), but have yet to be widely applied in this capacity to sex estimation. When compared to traditional statistical methods using osteometric data, ML approaches usually outperform, but only slightly and faired equally when using ordinal data. For example, Feldesman (2002) found

DFA and classification trees to perform equally well for morphometric hominin data. Du Jardin, Ponsaillé, Alunni-Perret, and Quatrehomme (2009) tested the classification accuracy of femur metrics using more traditional univariate, DFA, and LR as compared to ANN. ANN resulted in the highest classification accuracy (3.9% higher) and significantly less sex bias (i.e., classification accuracy differences between males and females). Likewise, Navega, Vicente, Vieira, and Cunha (2015) also demonstrated that ML methods, specifically decision trees and ANN, outperformed DFA and LR methods for metric sex estimation. In contrast, Curate et al. (2017) examined measurements of the femur using cross-validated DFA (88.4%), LR (88.4%), and SVM (89.1%) and found comparable accuracy rates among all three methods.

A review of all these statistical approaches leads us to the questions: (1) what statistical approaches should we be using to estimate sex from unknown remains? (2) Do specific data types (ordinal, continuous, etc.) limit or preclude the use of certain statistical approaches?

"Illegal statisticizing": The measurement statistics controversy

Stevens (1946) introduced the Theory of Admissible Statistics to outline measurement levels and permissible statistics, whereby nominal and ordinal scales should rely on nonparametric tests, while ratio and interval data should rely on parametric procedures. Stevens (1946, p. 679) suggested that "illegal statisticizing," or the application of statistical methods using data that they were not designed for, "can invoke a kind of pragmatic sanction: In numerous instances it leads to fruitful results." Unfortunately, nearly all researchers have since ignored Stevens' point about being pragmatic, especially in biological anthropology. By using acceptable statistical methods for each data type presented, researchers have ensured that the results from such analyses are valid and not an artifact of the statistical method employed (Scholten & Borsboom, 2009). These "rules" have become dogma when applied to statistics in biological anthropology and have largely gone unchallenged, although rarely do research articles address or test the assumptions and their subsequent results.

Each of the statistical methods discussed above has both strengths and limitations in actual application—in our case, sex estimation in unknown individuals. Walker (2008, p. 44) argues that "the test of the efficacy of a specific discriminant procedure in this context is not how well the data fit the assumptions of the technique, but how well the procedure solves the classification problem at hand." Hefner et al. (2014, p. 584) further suggest that "while classification statistics play a vital role in decedent identification in many, if not all, forensic anthropology laboratories, the foundational assumptions behind the statistics are often left unconsidered." This begs the questions: what statistical methods should we be using and do the statistical assumptions matter if we are not testing and reporting those assumptions?

Traditional statistical methods rely on a series of assumptions that must be tested prior to the application of a particular method with a particular data set. Theoretical assumptions were necessary in order to facilitate mathematical processing in the precomputer era, when calculations were performed with a slide rule. Times have changed. ML methods have been easier to implement, thanks to advances in computer software and hardware. Sex estimation is a real-world problem, in that it attempts to aid in the identification of an unknown person and has practical needs beyond the theoretical statistics. The research presented here enters biological anthropology into the *Measurement-Statistics Controversy*, which explores the practicality of bending the rules to improve higher classification accuracy in our case of sex estimation (Stevens, 1946). This controversy essentially focuses on Walker's comment above, or the fact that a correct classification should be more important on a practical level than abiding by statistical assumptions on a theoretical level. Is disregarding statistical assumptions necessary and fruitful for sex estimation (i.e., gives us a better classification), or should we stick to methods most appropriate for the data? We would argue, yes! This question parallels the use of ML methods, which can often produce higher classification accuracies with few assumptions, versus traditional statistical methods with their more numerous and explicit assumptions. Violating the rules when using traditional statistical methods may affect classification accuracies, but certainly affects the estimation of results such as overall significance, and posterior and typicality probabilities: If the rules are broken, these results can be quite deceiving; if the rules are bent, there may be little practical difference; if you want to avoid hard-and-fast rules, use ML methods. The key to choosing which classification method to use depends on the classification accuracy estimated from an independent holdout sample, which should represent the accuracy of the method when applied to a new case. Logically, the method with the highest accuracy and lowest bias when applied to holdout samples is the best method to use when estimating the sex of an unknown individual.

Testing the *Measurement-Statistics Controversy* as related to sex estimation

SPSS Modeler 18.1.1 was used to explore various classification methods for the Walker (2008) and Klales et al. (2012) ordinal trait scores contained within the MorphoPASSE Program Database ($n = 2366$) (see Chapter 16 of this volume). Statistical approaches specifically designed for ordinal data, such as LR, were compared to statistical methods in which the assumptions were violated by the data type, for example, LDFA, *k*NN, and Bayesian networks. Lastly, supervised ML methods were tested including ANN, SVM, and multiple DT variants: classification and regression trees (C&R), Quick, Unbiased, Efficient Statistical Tree (QUEST), Chi-squared Automatic Interaction Detection (CHAID), C5.0 trees, Tree-AS, and Random Forest. Gradient boosting algorithms (XGBoost) with linear and tree models as bases were also tested. XGBoost algorithms iteratively learn which of the predictor variables are weak and then add them back at the end to the final set of strong predictor variables using Python. Once the data were entered into the SPSS Modeler Stream, the automated modeling auto-classifier node was used to compare and contrast all models applicable to binary outcomes: sex (female/male). The models generated by the node were ranked from best to worst based on performance (i.e., classification) for the predictor variables selected (i.e., ordinal trait scores of the skull and innominate). The auto-classified node was run using default settings for the five traits of the skull (nuchal crest, glabella, mastoid process, supra-orbital margin, and mental eminence) and the three traits of the innominate (ventral arc, subpubic contour, medial aspect of the ischio-pubic ramus). Classification accuracy and sex bias in classification were explored to address *the Measurement-Statistics Controversy* and put forth suggestions for future analytical approaches to sex estimation.

For the skull, nearly every statistical approach outperformed LR (62.4%) with the exception of ANN (Table 1). Classifications were the highest using DT models, which made up seven of the eight highest classifying methods. XGBoost Tree (81.2%), RFM (79.8%), and C5.0 (79.2) were the three best performing statistical approaches. DFA ranked fourth best at 78.1% and was comparable to the top three decision tree approaches. With each of these top eight methods, males classified better than females. Sex bias was reduced using SVM, *k*NN, Bayesian and ANN, and LR, although classification was much lower with most of these approaches. In the innominate, again nearly

Table 1 Classification accuracy for the five skull traits using various statistical approaches.

Statistical approach	**Classification accuracy (%)**
XGBoost Tree	81.2
Random Trees	79.8
C5	79.2
Discriminant	78.1
Tree-AS	78.1
CHAID	77.9
Quest	77.1
C&R Tree	77.0
LSVM	75.4
XGBoost Linear	74.4
kNN Algorithm	63.3
Bayesian Network	62.8
SVM	62.5
Logistic Regression	62.4
Neural Net	60.1

Table 2 Classification accuracy for the three pelvis traits using various statistical approaches.

Statistical approach	Classification accuracy (%)
XGBoost Tree	88.2
Discriminant	87.5
C5	87.2
CHAID	86.9
Quest	86.7
C&R Tree	86.7
Random Tree	85.3
Tree-AS	85.5
LSVM	85.5
XGBoost Linear	79.6
SVM	74.8
Bayesian Network	74.7
Logistic Regression	74.7
Neural Net	74.7
kNN Algorithm	74.2

every statistical approach outperformed LR with the exception of ANN and *k*NN (Table 2). XGBoost Tree again performed the best (88.2%), followed by DFA (87.5%). Decision trees again made up seven of the top eight methods, and again males classified better than females. Sex bias was significantly reduced using SVM, LR, Bayesian and ANN, and *k*NN.

A further test of a theoretical statistics approach to analyzing data from the Klales et al. (2012) scoring method was provided by Konigsberg and Frankenberg (2019). They demonstrated high accuracies in estimating sex for an independent data set as well as a in a highly imbalanced simulated data set (975 males, 25 females) using ordinal probit. But a ML method, Random GLM (Song, Langfelder, & Horvath, 2013), performed just as well as the more sophisticated method they outlined, using far simpler coding (Klales, Ousley, & Vollner, 2019).

Results indicate that, in some instances, statistical approaches designed for other data types (i.e., LDFA) outperformed the methods designed specifically for ordinal data (i.e., LR). This indicates that the practical goals of statistical methods may be more relevant than theoretical dogma regarding the usage of data types when applying equation-based approaches like DFA and LR, which ultimately supports the notion of bending the rules as described in the *Measurement-Statistics Controversy*. This is especially true for classification methods in which classification accuracy is the overarching practical criterion (Velleman & Wilkinson, 1993; Walker, 2008). Results of this study also overwhelmingly indicate that newer ML models can best accomplish high classification rates and low sex bias using a variety of data types and may eventually replace the use of more traditional approaches.

Conclusion

The goals of this chapter were to briefly present an historical review of the most commonly used statistical approaches for sex estimation from human skeletal remains while also attempting to shed light on the *Measurement-Statistics Controversy* and where classification for the purpose of sex estimation falls within this controversy. Moving forward with sex estimation, we need to delineate the criteria for using various methods and how to best interpret if a method is generating valid and reliable results. A first step toward this process is requiring future methods to include universally understandable definitions of traits and/or measurements with illustrations, using validated statistical methodologies with intra- and inter-observer error rates. Further, the generation of published standards for method creation and application in biological anthropology would assist in the adoption of such criteria. At present, these criteria are not entirely clear in some of the newer supervised ML options, and we need them to be able to calculate the posterior probabilities of new cases. In the case of newer supervised ML models, testing should utilize independent validation samples to ensure broad applicability. This chapter has also shown that bending the rules for practical results is possible for accurate sex estimation; however, tests of each methodological assumption should be presented in research moving forward. As the field of biological anthropology evolves, there will likely be a move away from traditional statistical approaches like LR and DFA to more advanced and more appropriate supervised ML models, which, as this chapter has shown, perform equally as well if not better than more traditional approaches to sex estimation (cf. Klales et al., 2019).

References

Adams, D. C., Rohlf, J., & Slice, D. E. (2004). Geometric morphometrics: Ten years of progress following the 'revolution'. *Italian Journal of Zoology*, *71*, 5–16.

Aha, D. W. (1997). Lazy learning. *Artificial Intelligence Review*, *11*, 7–10.

Anastasiou, E., & Chamberlain, A. T. (2012). The sexual dimorphism of the sacro-iliac joint: An investigation using geometric morphometric techniques. *Journal of Forensic Sciences*, *58*(s1), S126–S134.

Atkinson, L. (1988). The measurement-statistics controversy: Factor analysis and subinterval data. *Bulletin of the Psychonomic Society*, *26*, 361–364.

Berry, K. J., Johnston, J. E., & Mielke, P. W. (2014). *A chronicle of permutation statistical methods: 1920–2000, and beyond*. New York, NY: Springer.

Boldsen, J. L., Milner, G. R., & Hylleberg, R. (2002). *ADBOU age estimation software. Version 2.0*.

Bruzek, J. (2002). A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology*, *117*, 157–168.

Brzobohatá, H., Krajíček, V., Velemínský, P., Poláček, L., & Velemínská, J. (2014). The shape variability of human tibial epiphyses in an early medieval Great Moravian population (9th–10th century AD): A geometric morphometric assessment. *Anthropologischer Anzeiger*, *71*, 219–236.

Bytheway, J., & Ross, A. (2010). A geometric morphometric approach to sex determination of the human adult os coxa. *Journal of Forensic Sciences*, *55*(4), 859–864.

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Points of significance: Statistics versus machine learning. *Nature Methods*, *15*, 233–234.

Chen, Y. C. (2017). A tutorial on kernel density estimation and recent advances. *Biostatistics and Epidemiology*, *1*, 161–187.

Curate, F., Umbelino, C., Perinha, A., Noguiera, C., Silva, A. M., & Cunha, E. (2017). Sex determination from the femur in Portuguese populations with classical and machine-learning classifiers. *Journal of Forensic and Legal Medicine*, *52*, 71–81.

Darroch, J. N., & Mosimann, J. E. (1985). Canonical and principal components of shape. *Biometrika*, *72*, 241–252.

Daubert vs. Merrell Dow Pharmaceuticals. (1993). US Supreme Court 09.U.S.579,113S.Ct.2786, 125L. Ed.2d 469.

Dirkmaat, D. C., & Cabo, L. C. (2012). Forensic anthropology: Embracing the new paradigm. In *Companion to forensic anthropology* (pp. 1–40). Malden, MA: Wiley-Blackwell.

Donnelly, S. M., Hens, S. M., Rogers, N. L., & Schneider, K. L. (1998). Technical note: A blind test of mandibular ramus flexure as a morphological indicator of sexual dimorphism in the human skeleton. *American Journal of Physical Anthropology*, *107*, 363–366.

Du Jardin, P., Ponsaillé, J., Alunni-Perret, V., & Quatrehomme, G. (2009). A comparison between neural network and other metric methods to determine sex from the upper femur in a modern French population. *Forensic Science International*, *192*, 1–3.

Feldesman, M. (2002). Classification trees as an alternative to linear discriminant analysis. *American Journal of Physical Anthropology*, *119*(3), 257–275.

Franklin, D., Freedman, L., Milne, N., & Oxnard, C. (2006). A geometric morphometric study of sexual dimorphism in the crania of indigenous southern Africans. *South African Journal of Science*, *102*, 229–238.

Galdames, I. C. S., Valenzuela, J. S. P., Quezada, N. A. S., Contreras, C. E. C., Rivas, J. A. H., & López, M. C. (2008). Ortopantomographic blind test of mandibular ramus flexure as a morphological indicator of sex in Chilean young adults. *International Journal of Morphology*, *26*(1), 89–92.

Giles, E., & Elliot, O. (1963). Sex determination by discriminant function analysis of crania. *American Journal of Physical Anthropology*, *2*, 53–68.

Hefner, J. T. (2009). Cranial nonmetric variation and estimating ancestry. *Journal of Forensic Sciences*, *54*(5), 985–995.

Hefner, J. T., & Ousley, S. D. (2014). Statistical classification methods for estimating ancestry using morphoscopic traits. *Journal of Forensic Sciences*, *59*, 883–890.

Hefner, J. T., Spradley, M. K., & Anderson, B. E. (2014). Ancestry assessment using random forest modeling. *Journal of Forensic Sciences*, *59*, 583–598.

Jantz, R. L., & Ousley, S. D. (2005). *FORDISC 3: Computerized forensic discriminant functions. Version 3.0.* Knoxville, TN: The University of Tennessee, Knoxville.

Kimmerle, E. H., Jantz, R. L., Konigsberg, L. W., & Baraybar, J. P. (2008). Skeletal estimation and identification in American and East European populations. *Journal of Forensic Sciences*, *53*(3), 524–532.

Klales, A. R. (2013). Current practices in forensic anthropology for sex estimation in unidentified, adult individuals. In *Proceedings of the 65th annual scientific meeting of the American Academy of Forensic Sciences, 19*. Colorado Springs, CO: American Academy of Forensic Sciences.

Klales, A. R., Ousley, S. D., & Vollner, J. M. (2012). A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *American Journal of Physical Anthropology*, *149*, 104–114.

Klales, A. R., Ousley, S. D., & Vollner, J. M. (2019). Response to multivariate ordinal probit analysis in the skeletal assessment of sex (Konigsberg and Frankenberg). *American Journal of Physical Anthropology*, *169*, 388–389.

Konigsberg, L. W., & Frankenberg, S. (2019). Multivariate ordinal probit analysis in the skeletal assessment of sex. *American Journal of Physical Anthropology*, *169*, 185–387.

Konigsberg, L. W., & Hens, S. M. (1998). Use of ordinal categorical variables in skeletal assessment of sex from the cranium. *American Journal of Physical Anthropology*, *107*, 97–112.

Krishan, K., Chatterjee, P. M., Kanchan, T., Kaur, S., Baryah, N., & Singh, R. K. (2016). A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. *Forensic Science International*, *261*, 165.

Kustár, Á., Gerendás, Z., Kalina, I., Fazekas, F., Vári, B., Honti, S. z., et al. (2013). FACE-R—3D skull and face database for virtual anthropology research. *Annales Historico-naturales Musei Nationalis Hungarici*, *105*, 313–319.

Lesciotto, K. M. (2015). The impact of Daubert on the admissibility of forensic anthropology expert testimony. *Journal of Forensic Sciences*, *60*, 549–555.

Loth, S. R., & Henneberg, M. (1996). Mandibular ramus flexure: New morphologic indicator of sexual dimorphism in the human skeleton. *American Journal of Physical Anthropology*, *99*, 473–486.

Milner, G., & Boldsen, J. (2012). Transition analysis: A validation study with known-age modern American skeletons. *American Journal of Physical Anthropology*, *148*(1), 98–110.

Mizell, S. M., Long, H., & Klales, A. R. (2019). The application of the Optimized Summed Scored Attributes (OSSA) method to sex estimation in Whites and Blacks. In *Proceedings of the 71st annual scientific meeting of the American Academy of Forensic Sciences, 25*. Colorado Springs, CO: American Academy of Forensic Sciences.

Murphy, R. E., & Garvin, H. M. (2018). A morphometric outline analysis of ancestry and sex differences in cranial shape. *Journal of Forensic Sciences*, *63*(4), 1001–1009.

Navega, D., Vicente, R., Vieira, D. N., Ross, A. H., & Cunha, E. (2015). Sex estimation from the tarsal bones in a Portuguese sample: A machine learning approach. *International Journal of Legal Medicine*, *129*(3), 651–659.

Ousley, S. D., & Kenyhercz, M. (2013). Geometric morphometrics and statistical classification: Size matters. *American Journal of Physical Anthropology*, *150*(S56), 1–301.

Perlaza, N. A. (2014). Sex determination from the frontal bone: A geometric morphometric study. *Journal of Forensic Sciences*, *59*(5), 1330–1332.

Phenice, T. W. (1969). A newly developed visual method of sexing the os pubis. *American Journal of Physical Anthropology*, *30*, 297–301.

Pohar, M., Blas, M., & Turk, S. (2004). Comparison of logistic regression and linear discriminant analysis: A simulation study. *Advances in Methodology and Statistics*, *1*(1), 143–161.

Rogers, T. L., & Saunders, S. (1994). Accuracy of sex determination using morphological traits of the human pelvis. *Journal of Forensic Sciences*, *39*, 1047–1056.

Ross, A. H., Slice, D. E., & Williams, S. E. (2010). *Geometric morphometric tools for the classification of human skulls*. National Institute of Justice Grant. Award Number: 2005-MU-BX-K078.

Scholten, A. Z., & Borsboom, D. (2009). A reanalysis of Lord's statistical treatment of football numbers. *Journal of Mathematical Psychology*, *53*(2), 69–75.

Slice, D. E., & Ross, A. (2009). *3D-ID: Geometric morphometric classification of crania for forensic scientists*. *Version 1.0*.

Song, L., Langfelder, P., & Horvath, S. (2013). Random generalized linear model: A highly accurate and interpretable ensemble predictor. *BMC Bioinformatics*, *14*, 5.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680.

Stewart, T. D. (1942). Variation in the technique of measuring skulls. *Anthropology Briefs*, *2*, 1–6.

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Boston: Allyn and Bacon.

Tallman, S. D., & Go, M. C. (2018). Application of the optimized summed scored attributes to sex estimation in Asian crania. *Journal of Forensic Sciences*, *63*(3), 809–814.

Velleman, P. F., & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *The American Statistician*, *47*, 65–72.

Vollner, J. M., Klales, A. R., & Ousley, S. D. (2009). Sexing of the human innominate using non-metric traits and statistical analysis. *American Journal of Physical Anthropology*, *138*, 263–264.

Walker, P. L. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, *136*, 39–50.

Washburn, S. L. (1948). Sex differences in the public bone. *American Journal of Physical Anthropology*, *6*, 199–208.

Williams, B. A., & Rogers, T. L. (2006). Evaluating the accuracy and precision of cranial morphological traits for sex determination. *Journal of Forensic Sciences*, *51*(4), 729–735.

CHAPTER 14

Subadult sex estimation and KidStats

Kyra E. Stull[a,b], Laura E. Cirillo[a], Stephanie J. Cole[a], Cortney N. Hulse[a]
[a]Department of Anthropology, University of Nevada-Reno, Reno, NV, United States
[b]Department of Anatomy, University of Pretoria, Pretoria, South Africa

Introduction

It has been demonstrated throughout the literature that childhood stature, weight, body composition, and developmental trajectory, velocity, and tempo differ between the sexes (Arfai et al., 2002; Bogin, 1999; Cameron, Tanner, & Whitehouse, 1982; Eveleth, 1978; Freedman, Khan, Serdula, Ogden, & Dietz, 2006; Hutt, 1972a, 1972b; Matthews et al., 2018; Ounsted & Taylor, 1972; Schiessl, Frost, & Jee, 1998; Stinson, 1985; Tanner, 1989; Tanner, Hughes, & Whitehouse, 1981; Wells, 2007). These differences have even prompted the development of sex-specific growth curves. It should not be surprising then that these same sex differences could be reflected in the subadult skeleton prior to puberty. Yet, subadult sex estimation is routinely discouraged in forensic and bioarchaeological application for the commonly stated reason that sexually dimorphic differences do not exist in the skeleton prior to puberty (SWGANTH, 2013).

There is an apparent disconnect between biological anthropologists and other researchers who have documented and quantified differences between the sexes. In order to identify these differences in the skeleton and subsequently develop models, there must be sufficient samples for data collection, which is the prevailing obstacle in any subadult research. Most subadult skeletal collections lack an appropriate sample size for developing methods in anthropology, both in number and in adequate reflection of the population. Sample size issues are felt at a greater magnitude in subadult research because an appropriate sample size is required for all chronological ages that are being included in the study. For example, if a study includes individuals from birth to 5 years of age, one will need to have adequate samples of individuals within each chronological year in order to capture all the changes that occur during that important developmental period.

To compensate for small sample sizes, some authors will pool ages (e.g., Klales & Burns, 2017; Viciano, López-Lázaro, & Alemán, 2013; Wilson, Cardoso, & Humphrey, 2011). Pooled age ranges may mask differences and underrepresent developmental changes. Further, too small a sample for a large age range provides no substantial information for each subset of the data. If researchers are adamant to pool ages, the age subsets should be established on a trait-by-trait basis rather than arbitrary age cohorts (Wilson & Humphrey, 2017). In other words, divisions should be based upon examination of the

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00014-6

ontogenetic trajectory for each indicator under study and not based on a priori assumptions. If the same age subsets are applied to all sex indicators being explored, there is an implicit assumption that all indicators follow the same rate and onset of sexual dimorphism, and have the same selective pressures or proximate and ultimate mechanisms. Adult sex indicators do not follow the same path for expression, so it is logical to assume that holds true for subadult expression as well.

Outside of the limitations of too few substantial collections, there are other reasons why the available collections are not appropriate for use in modern forensics. Several subadult skeletal collections arose from the exhumation of historic cemeteries (e.g., Granada, Lisbon, Spitalfields). Historic cemetery populations, or any sample comprising individuals who lived in a noncontemporaneous period, cannot be assumed to present with variation that is representative of a modern population. Secular trends in childhood obesity and onset of maturation, as well as growth differences resulting from disparate environments, are just a few of the reasons as to why this could be problematic (Klepinger, 2001).

Without subadult samples available, researchers are very limited in what new methodologies they can develop to supplement the subadult biological profile. For almost a decade, researchers have been attempting to create subadult collections, but rather than being built on skeletal remains, they instead use medical images such as conventional radiography, magnetic resonance images (MRI), ultrasound, and computed tomography (CT) scans. The first large, freely available collection of subadult radiographs is the Pediatric Radiology Interactive Atlas, or PATRICIA (Ousley, 2013, NIJ 2008-DN-BX-K152). PATRICIA comprises 44,220 radiographic images of 9709 individuals from both clinical and morgue settings. Since then, researchers have developed radiographic databases that include both adults and subadults, and others are continuing to develop databases that include individuals from a wide array of geographic and economic diversities. One database is currently being created from full-body CT scans of children from the United States, France, the Netherlands, Brazil, and Taiwan (Stock, Stull, Garvin, & Klales, 2016, NIJ Awards 2015-DN-BX-K409, 2017-DN-BX-0144). The database will also include radiographic images from Angola and South Africa and raw data from Colombian children. The creation of these large collections facilitate subadult sex estimation research, and subadult research more broadly.

The goal of this chapter is to provide information regarding the intricacies of sexual differentiation and the expression of sexual dimorphism and to introduce the publications that have attempted to disprove the consistent belief that sex estimation is not possible in the subadult skeleton. The last third of the chapter is dedicated to highlighting the biological complexities that can impede high accuracies in subadult sex estimation as well as several methodological complexities that may hinder success. Some of the topics include secular trends, population variation, measurement theory, and imbalanced classes. While there may be some discouraging results along the way, we hope the reader can persevere and be encouraged to continue research and ultimately transform the field's thoughts on subadult sex estimation. The lack of usable methodologies is not demonstrative of potential outcomes. If anything, this chapter emphasizes the available research niche of subadult sex estimation that could substantively impact the field of forensic anthropology and bioarchaeology.

Complex underpinnings of sexual differentiation and sexual dimorphism

Scientists in many fields have sought to understand the proximate and ultimate mechanisms that inform the expression of dimorphic phenotypes from the nearly identical genomes of the sexes (Cox, Stenquist, & Calsbeek, 2009; Rhen, 2007). Most biological anthropologists believe sexually dimorphic features result from puberty. However, puberty is not one single life event; rather it is one stage of reproductive life that begins during embryonic development. As such, there are many dependencies that inform the final phenotype prior to puberty (Lee & Styne, 2013; Styne & Grumbach, 2011).

Hormone regulation is considered one of the greatest mechanisms that affect sex differences in phenotypic traits, especially since most genes that inform sexual dimorphism are not sex-linked (Badyaev, 2002; Bernstein, 2010). The apparent sexual (size) dimorphism in humans is primarily driven by gonadotropins, namely luteinizing hormone (LH) and follicle-stimulating hormone (FSH), which stimulate the gonads (i.e., testes or ovaries). These two hormones are secreted in both sexes, but the roles each play are dependent on the sex of the individual. For example, in males, FSH is essential for sperm production and LH is responsible for testosterone synthesis and secretion. In females, FSH is responsible for estrogen production and maturation of follicles and LH is fairly inactive until menarche when it is then responsible for the ovulation of mature follicles.

The first stages of differentiation commence around 7–9 weeks of gestation with the activation of the sex-determining region of the Y-chromosome (SRY) and is complete at approximately 20 weeks of gestation. While the *SRY* gene is recognized as being the catalyst for male sex differentiation, its downstream target, the *SOX9* gene, is the element that orchestrates Sertoli cell differentiation. Sertoli cells support germ cells in the testes and, therefore, direct testis morphogenesis (Kashimada & Koopman, 2010; McClelland, Bowles, & Koopman, 2012). A cascade of events ensues to develop a male or female embryo. The hormonal surge experienced by males in utero has been linked to social and cognitive differences (Alexander & Wilcox, 2012) and may also be responsible for sexually dimorphic characteristics, such as increased bone and muscle mass, higher birth length and weight, and larger head circumference in males compared to females (Largo, Walli, Duc, Fanconi, & Prader, 1980; Lubchenco, Hansman, & Boyd, 1966; Lubchenco, Hansman, Dressler, & Boyd, 1963; Olsen, Groveman, Lawson, Clark, & Zemel, 2010; Thomas, Peabody, Tunier, & Clark, 2000).

The surge in hormone levels in early development in utero is considered a mini-puberty by some researchers because median levels of hormones are comparable to median levels during puberty (e.g., Aksglaede, Juul, Leffers, Skakkebæk, & Andersson, 2006; Grumbach, 2002; Mann & Fraser, 1996). The activation of the hypothalamus-pituitary-gonad (HPG) axis, resulting in high gonadotropin levels, is present in males and females during the first year of life. Males experience increased levels of LH until about 6 months of age, and females have increased levels of FSH until approximately 12–24 months. In response to heightened hormone levels, males and females secrete testosterone or estradiol, respectively (Grumbach, 2002; Lanciotti, Cofini, Leonardi, Penta, & Esposito, 2018). After the first year of life, gonadotropins rapidly decrease and remain relatively hypoactive until the HPG axis is re-activated as the child enters puberty (Aksglaede et al., 2006).

Sexual differences we recognize in adults are largely and indirectly controlled by sex steroids as they inform the induction of developmental timings and the resulting expression of sexual dimorphism. Just prior to pubertal onset, there is an increase in gonadal steroids following an increase in the amplitude of gonadotropin pulses. Gonadotropins differentiate the sexes by organizing distinct growth rates, growth durations, and reproductive functions (Badyaev, 2002; Chou, Iwasa, & Nakazawa, 2016; Shea, 1992; Stulp & Barrett, 2016). A developmental perspective is especially important for understanding variation in body size and the differences expressed by both sexes (Bernstein, 2010; Leigh & Shea, 1995). Sexual size dimorphism (SSD) has been shown to slowly increase up to approximately 8 years of age, and then decrease until early teenage years. This decrease in size differences is due to the differences in age of onset of pubertal growth spurt; specifically, females increase in size earlier than males, resulting in similar body sizes for a short period. Once males enter their pubertal growth spurt, the increase in SSD is observed once again and increases with age (Nikitovic & Bogin, 2014). Sexual dimorphism in these growth patterns and developmental processes leads to adult SSD, while gonadal steroids lead to the expression of secondary sexual characteristics commonly associated with puberty.

While it is difficult to believe that major differences would exist in subadult skeletal elements of males and females, let alone in infant or fetal skeletons, research states that by 6 h postfertilization, males grow at five times the rate of females (Burgoyne et al., 1995; Mittwoch, 1993; Ray, Conaghan, Winston, & Handyside, 1995). The differential growth is linked to sexual differentiation and the activation of genes on the Y-chromosome that are the catalysts for differentiation. Interestingly, this faster initial growth has been linked to the female buffering hypothesis (Badyaev, 2002), which states that males have increased susceptibility to suboptimal environmental conditions during growth and development. The female buffering hypothesis is, therefore, linked to the expression of sexual dimorphism in a population. Specifically, in less-than-favorable conditions, sexual dimorphism is reduced because of males' increased susceptibility to environmental factors, and similarly, males have increased mortality rates (Nikitovic & Bogin, 2014). This example elucidates that some aspects of our phenotype are informed early in the developmental process and are not just a consequence of puberty.

Why does the pedagogical mantra exist?

Despite the general recommendation to avoid subadult sex estimation in biological and forensic anthropological practice for individuals <12 years (SWGANTH, 2013), some authors pursue the exploration of sexual dimorphism in the subadult skeleton in hopes of identifying a method to accurately estimate sex. The research that has been conducted thus far on each major anatomical area is discussed below. The many studies have been employed across the entire age range, as there are few opportunities for inclusion/exclusion criteria, and instead researchers use available specimens. In the discussion below, we do not always specify the age ranges included in the studies. This section introduces the trends in methods for subadult sex estimation based on the crania and postcrania and summarizes both the criticisms that have enforced the perpetuation of current thoughts and the methods that are shaping current practice. Table 1 provides a summary of all the information discussed below and includes important criteria for many of the cited studies.

Crania

The vast majority of methods for estimating sex from the subadult skull focus on observations of the teeth and mandible. Protrusion of the chin, shape of the anterior dental arcade, and eversion of gonion have been presented as morphologically distinct between males and females and used as subadult sex indicators. These traits had high accuracies for classifying males (74%–94.1%), but often failed to accurately classify females (Schutkowski, 1993). The shape of the inferior border of mandibular symphysis and the outline of mandibular body have also exhibited differences between the sexes with a classification accuracy around 81%, with a bias in favor of male classification (Loth & Henneberg, 2001). Subsequent validation studies of both methods failed to achieve the reported accuracies (Scheuer, 2002).

Some authors have modified existing methods used in adult sex estimation in their exploration of subadults. Molleson et al. (1998) used a combination of mandibular and splanchnocranium features (e.g., mandibular angle, supraorbital margin, mentum) that were adopted from Acsadi and Nemeskeri (1970), but eliminated "hyperfeminine" and "hypermasculine" expressions. The method correctly estimated sex in 78% of the small test sample (Molleson et al., 1998). Despite modifying the degree of expression expected in subadults, these methods still rely on the same sexually dimorphic traits observed in adults. This problematic projection of adult expectations onto subadult elements is a common theme in the literature.

Table 1 Summary table of subadult sex estimation studies including their sample size, anatomical focus, and their results. The "external validation result" column provides some validation numbers, but also some references, depending on the original study.

Reference	Sample size	Skeletal region	Age ranges	*M* accuracy	*F* accuracy	External validation result
Andras and Stock (2018)	56	Ilium	1 day to 1 year	59.1%–73.53%		
Bilfeld et al. (2015)	188	Pubis	1–18 years			
Black (1978)	133	Deciduous dentition	Unknown	66.7%–72.5%	57.8%–68.8%	
Cardoso (2008)	156	Dentition	1.17–15 years; 20–56 years	25%–100%	30%–100%	
Cardoso (2010)	46	Dentition	Birth to 10 years	33.3%–75%[a]		46.2%–60%
Choi and Trotter (1970)	115	Long bones (ratios)	Fetal	72%[a]		
De Vito and Saunders, 1990	162	Dentition	3–4 years and 16 years	76%–90%[a]		
Estévez Campo, López-Lázaro, López-Morago Rodríguez, Alemán Aguilera, and Botella López (2018)	83	Pelvis	Birth to 1 year	48.3% (pubis), 54.2% (ischium)[a]		
Garvin et al. (2019)	202	Pelvis	Birth to 10 years	60%–77%[a]		
Hassett, 2011	108	Canine		93.8%–95%	65%–87.5%	
Klales and Burns (2017)	334	Pubis	1.19–20.47 years	55.6%–100%	53.9%–97.2%	

López-Lázaro, Alemán, Viciano, and Irurita (2018)	68	Deciduous M1	Subadults	93.23%–100%	83.17%–87.5%	82.35–92.31%
Loth and Henneberg (2001)	62	Mandible	Birth to 19 years	81%[a]		64% (Scheuer, 2002)
Lund and Mörnstad (1999)	58	Dentition	14–38 years			
Mittler and Sheridan (1992)	58	Ilium	Birth to 18 years	85.3%	58.3%	
Molleson, Cruse, and Mays (1998)	20	Skull	1–14 years	78%[a]		
Schutkowski (1993)	61	Mandible and illium	Birth to 5 years	70%–90%[a]		Sutter (2003), Cardoso and Saunders (2008), Loth (1996), and Irurita Olivares and Alemán Aguilera (2016)
Stull and Godde (2013)	85	Long bones	Birth to 1 year	96.7% (femur); 88.6% (humerus)		
Stull, L'Abbe, and Ousley (2017)	1310	Long bones	Birth to 12 years	74%–95%	72%–90%	
Viciano et al. (2013)	269	Dentition	Infant to adult	78.1%–93.1%[a]		
Weaver (1980)	153	Ilium	Fetal to 6 months	73.1%–91.7%	43.5%–75.0%	Hunt (1990) and Mittler and Sheridan (1992)
Wilson, MacLeod, and Humphrey (2008)	25	Ilium (GSN)	0–7.88 years	100%	88%	Wilson et al. (2011)
Zadzinska, Karasinska, Jedrychowska-Danska, Watala, and Witas (2008)	113	Deciduous dentition	Subadult	69%	88%	

[a] Indicates that the study collapsed accuracy between males and females to report overall accuracy.

Dental variation between the sexes is an ideal medium for both bioarchaeological and forensic subadult sex estimation research due to the high preservability of teeth. Although there are disagreements as to which degree dentition can be used to estimate the sex of subadults, numerous studies have identified the first and second maxillary and mandibular molars as the most sexually dimorphic deciduous teeth (Black, 1978; Cardoso, 2010; López-Lázaro et al., 2018; Margetts & Brown, 1978; Viciano et al., 2013; Zadzinska et al., 2008). This is in contrast to permanent dentition, where the maxillary and mandibular canines present with the greatest degree of sexual dimorphism (Cardoso, 2008; Hassett, 2011; Lund & Mörnstad, 1999; Moorrees, Fanning, & Hunt, 1963). Accuracies ranged between 63.9% and 90.5% in studies using mesiodial and buccolingual measurements from deciduous tooth crowns in discriminant function analysis (DFA) (Black, 1978; De Vito & Saunders, 1990). When measurements from the permanent first molars were included, accuracies increased and sex bias shifted in favor of female classification (~11%). Classification accuracies approached rates seen with the permanent dentition (Cardoso, 2008; De Vito & Saunders, 1990; Garn, Cole, Wainwright, & Guire, 1977; Hassett, 2011). Viciano et al. (2013) reported correct sex assignments between 78.1% and 93.1% when using the first and second deciduous molar and permanent canine, and between 79.4% and 92.6% when using permanent teeth. The accuracy values were substantially higher than those in previously published studies, likely due to the use of a combined dentition sample, inclusion of cervical measurements, and multivariate models. These findings demonstrate high rates of sexual dimorphism in dental diameters, which contrasts what some authors (e.g., Cardoso, 2010) have argued.

Postcrania

Subadult sex estimation studies based on postcranial elements generally concern the pelvis or long bones. Subadult sex estimation in the pelvis has been explored through metric and morphological approaches, and more recent studies are incorporating geometric morphometrics and more robust statistical analyses. The ilium is the most frequently examined pelvic element in subadult sex estimation, potentially because it is the largest bone of the ossa coxarum (Weaver, 1980). Despite some inconsistencies, most authors have identified significant shape and size differences with increased age (Mittler & Sheridan, 1992; Wilson et al., 2008, 2011; Bilfeld et al., 2013).

Auricular surface elevation has been viewed both as an indicator of sexual dimorphism and as a consequence of aging (Hunt, 1990; Weaver, 1980). Accuracies for both sexes improved in individuals >9 years of age, though females still only correctly classified slightly better than chance (58.3%). In contrast, male classification accuracy reached 85% (Mittler & Sheridan, 1992). Traits associated with the greater sciatic notch and iliac crest were established by Schutkowski (1993) and have since been tested by numerous authors. Males had a high classification accuracy when an acute notch was observed

(95%), but a much lower classification accuracy was produced when an obtuse notch was documented (71%). Additionally, the shape of the iliac crest was an overall poor indicator of subadult sex, with 81% of males exhibiting a marked S shape and only 62% of females exhibiting a faint S shape. Subsequent validation studies have yielded contradictory results and failed to achieve the accuracy rates reported in the original article (Cardoso & Saunders, 2008; Irurita Olivares & Alemán Aguilera, 2016; Sutter, 2003).

Sex estimation from the ilium using geometric morphometrics resulted in poor classification with regard to the iliac crest and auricular surface, with male accuracies far exceeding female accuracies (males = 82%–88%; females = 25%–38%). The measurements obtained from the greater sciatic notch resulted in much higher classification accuracies, with males and females achieving rates as high as 100% and 87.5%, respectively (Wilson et al., 2008). Consistent with the established trend, when the method was tested on a new sample, classification rates were significantly lower (Wilson et al., 2011). In the first year of life, males were significantly larger than females when exploring the ilium with interlandmark distances, Procrustes coordinates, and principal components, but the resulting classification accuracies via jackknifed linear DFA only ranged from 59% to 74% (Andras & Stock, 2018). When ilium outlines, greater sciatic notch measurements (length, depth, and angle), and indices of the pubis/ischial length of an older sample were examined, sexual dimorphism was only significant in components related to the ilium outline (Garvin et al., 2019). Only at 4 years of age did correct DFA classifications increase to 77.4% from 60.4%, indicating an age threshold for classification accuracy. In studies that included the pubis and ischium, there were no significant differences documented between the sexes (Estévez Campo et al., 2018; Garvin et al., 2019). Research published by Bilfeld et al. (2015) suggests sexually dimorphic differences in the pubic bone are not significant until individuals reach 13 years of age, although visible shape differences, though not significant, were observed in individuals as young as 9 years.

The long bones have received little attention regarding the development of sex estimation methods for subadults relative to other elements of the skeleton. Choi and Trotter (1970) and Stull and Godde (2013) used long bone lengths and breadths to investigate subadult sex differences using a sample of fetuses and infants, respectively. It is important to note that both studies had acceptable results, but used a limited number of measurements and narrow age ranges. To address these shortcomings, Stull et al. (2017) performed a follow-up study using 18 measurements from all six long bones on a larger sample. Multiple statistical analyses, including linear and flexible discriminant analysis and logistic regression, were employed using both single and multiple measurement models. Because the age range in this study was large (birth to 12 years), each model was run twice to include and exclude age as a covariate. Proximal and distal breadth measurements were found to be more important in model creation than length measurements, and models using multiple variables consistently gave higher classification accuracies compared to single variables. Flexible discriminant analysis gave the highest overall accuracies

(74%–93%) with low sex bias (∼3.5%); however, logistic regression models gave similar accuracies (72%–90%) with smaller standard errors, albeit with higher sex bias (∼9%). In both cases, sex bias was in favor of male classification. The inclusion of age was not found to consistently increase or decrease classification. The results of this study suggest that sex estimation from long bones can yield accuracies in subadults comparable to those obtained for adults (Stull et al., 2017).

Reconsideration of previous subadult sex research

The summary provided in Table 1 demonstrates many of the challenges that previous research has faced and supports reasoning for the consistency in the belief that subadult sex should not be estimated. Although there are a few original publications that practice methods that yield decent classification accuracies, the field has far to go before subadult sex estimation becomes part of standard practice.

The expectation that adult sex indicators would be present in subadults is one example of the problematic assumptions that currently underlie this area of research. Adult and subadult sex indicators are fundamentally different and, as such, should assume differences in traits and expression. Research designs founded in appropriate theoretical reasoning and measurement theory are crucial for direct interpretations of results (Houle, Pelabon, Wagner, & Hansen, 2011). Measurements must be linked to the theoretical and instrumental context from which they were derived, and subsequently, the inferences made from the measurements should reflect the underlying reality that we intend to capture (Houle et al., 2011). Measurement theory has not always been practiced in subadult sex estimation, as exemplified by research on the pelvis. The female pelvic basin follows a distinct growth trajectory that is slower and more constant compared to stature (Moerman, 1982). While the rest of skeletal maturation is heavily influenced by the adolescent growth spurt, the female pelvis will not reach full maturity until several years after adult height is achieved with the adolescent growth spurt (Bogin, 1999). In order to prioritize future reproductive success, this particular skeletal growth in females is given precedence over other growth, such as muscle mass and fat deposits (Stulp & Barrett, 2016). Greater success in classification may be achieved with the subadult pelvis if researchers considered the unique growth of the pelvis that is largely driven by reproduction. Based on our knowledge of sexual dimorphism in the pelvis, there is no logical reason to assume we would have success estimating the sex of subadults using adult sex indicators.

The pelvis is successful in adults, but the mechanisms directing that dimorphism should not theoretically direct dimorphism in subadults. For any of the pelvic sex indicators of adults, the application to subadults should first start with the ontogenetic appearance and earliest expression. Subadult sex indicators may not be recognized in the adult skeleton and, therefore, require developmental underpinnings and/or clinical or biomechanical literature to facilitate identification. It may not be that meaningful sex estimations are not possible for subadult skeletal remains, but that researchers have yet to collect data on appropriate elements that allow us to properly interpret the findings. Thus, measurement theory should be the foundation to any research design, especially in aspects of the biological profile that are more nuanced to identify and estimate than others.

Poor performance in validation studies may have more to do with two other aspects, namely the strength, or even replicability, of sex indicators and the modifications of traits, such that researchers are not testing the original study. Population variation has been blamed for the inconsistent findings, but a strong sex indicator should have little influence from population-level differences. Ideally, once a strong sex indicator has been identified, it could then be applied to numerous populations to explore its resilience to population or environmental impacts. Even so, reaching adult standards of expression are likely difficult to obtain. The sexually dimorphic difference is going to be less extreme in features of subadults than in features of adults. Although we may not currently be at a place to explicitly state an expected classification rate, we could consider the previously proposed 75% as the threshold when interpreting model performance until the field has developed to include more meaningful performance evaluations (DiGangi & Moore, 2012).

Factors impacting sex estimation

Secular trends

In human biology and anthropology, secular trends in children are generally discussed in terms of somatic and sexual maturity indicators. Numerous authors have documented secular trends in puberty for both males and females, as seen by lower ages at thelarche (breast development), menarche, and testicular volume (Chumlea, Schubert, & Roche, 2003; Herman-Giddens et al., 1997, 2012; Lee & Styne, 2013; Sørensen et al., 2012). Even though the onset of puberty is the result of a complex relationship between genetics and the environment, the secular trends associated with the onset of puberty are generally attributed to environmental changes since genetics are not drastically altered from one generation to the next (Chasiotis, Scheffer, Restemeier, & Keller, 1998; Danker-Hopfe, 1986; Euling, Selevan, Pescovitz, & Skakkebaek, 2008; Golub et al., 2008; Lee & Styne, 2013). Skeletal maturity is discussed with regard to the completion of epiphyseal fusion and not when sexually dimorphic features of the skeleton would be fully expressed. However, anthropologists should think of sexual dimorphism in relation to sexual maturity and not skeletal maturity. It is possible that if we are seeing evidence of secular trends for secondary sexual characteristics, we may also see evidence of sexually dimorphic features in younger individuals.

Secular change has been a constant influence on the timing and tempo of maturation and varies by country and socioeconomic status (SES) (Eveleth, 1978; Hauspie, Vercauteren, & Susanne, 1997; Tanner, 1992). The differences seen in the timing and tempo of maturation will affect the feasibility of predicting sex at certain ages. A close relationship is observable between the appearance and scoring of secondary sexual characteristics and bone maturity levels for both males and females (Harlan, Grillo, Cornoni-Huntley, & Leaverton, 1979; Harlan, Harlan, & Grillo, 1980; Herman-Giddens et al., 1997). Similarly, the age at take-off is directly correlated with the age at which that individual will reach puberty; menarche occurs, on average, 1 year after the age of peak height velocity, which is subsequent to take-off (Abbassi, 1998; Cole, 2003). However, the timing will always be related to secular trends and SES. When these relationships are taken into consideration, we might be able to understand the impact that secular change has on the appearance of secondary sexual characteristics, which will impact the forensic anthropologist's ability to estimate sex, specifically at a population level.

Expression of adult sex indicators in subadults

It is necessary in biological anthropology to explore the ontogenetic appearance of sexually dimorphic features, as this would inform practitioners of when they are able to confidently estimate sex. Using a culturally determined legal adult chronological age as the moment that one can accurately estimate sex is also problematic. For example, in the United States, 18 years is used as the arbitrary cut-off between adults and subadults without a link to any specific indicator of biological maturation and, therefore, has been problematic as a designated age threshold when applied to refugees/asylum-seekers. Additionally, because some features progress from a neutral state to a more robust state (e.g., cranium) and other features progress from a neutral state to a more gracile state (e.g., pelvis), one has to be aware of the underlying biological reasons for change prior to observing the feature. For example, in the case of a teenager with pelvic features indicative of female, a practitioner may be comfortable estimating sex because it is understood that pelvic morphology progresses to a gracile state in biologically female individuals. If the pelvis did not exhibit female-indicative features, one likely would not estimate sex because there would be no confidence at what age we would expect to see this change. This pattern holds true for any sex indicator until the forensic anthropologist is comfortable that expression should be complete at the age of the individual.

Because of these complexities in subadult sex estimation, some researchers began to move in the direction of exploring the onset of sexual dimorphism. Klales and Burns (2017) modified subpubic concavity stages and tested it on a radiographic sample of known sex-and-age individuals. The authors divided the sample into six age cohorts: Young Child Early (1.0–3.5 years), Young Child Late (3.6–6.5 years), Older Child Early (6.6–9.5 years), Older Child Late (9.6–12.5 years), Adolescent Early (12.6–15.5 years), and Adolescent Late (15.6–20.5 years). Logistic regression classification accuracies were determined for each age cohort. The accuracies for the three youngest cohorts ranged between 53.9% and 64.5%, while those for the three oldest cohorts ranged between 71.7% and 97.2%. More specifically, both Adolescent Early (12.6–15.5 years) and Adolescent Late (15.6–20.5 years) cohorts achieved accuracies exceeding 85%.

Males were more often correctly classified than females; this was especially true with regard to the two Young Child categories, where sex biases were found to exceed 80% in favor of male classification. Despite low classification accuracies for the younger cohorts, both Adolescent cohorts achieved classification rates approaching those observed in adults, which suggests that 12 years of age may be young enough to estimate sex.

Stock (2018) explored the expression of sexually dimorphic cranial traits in comparison to dental maturity, as gauged by alveolar eruption of the mandibular M3 with ¾ root completion following Moorrees et al. (1963) stages. Results indicated that the only cranial variable that matured in size earlier than when dentition reached maturity was the mental eminence. The expressions of nuchal crest and glabella were observed later than dental maturity and, in fact, were only completely expressed by some male individuals in the sample at age 22 or 23 years. This research provides good support that we can only confidently use adult sex indicators when we know the age at which sexually dimorphic features become fully expressed in both sexes.

Population variation and size differences

Researchers consistently interpret discrepancies between the results of validation studies and those of the original publication as population variation. Admittedly, the impact of population variation has not been fully assessed in subadult sex indicators, but sexual dimorphism has been demonstrated to outweigh population variation in adults (Kenyhercz, Klales, Stull, McCormick, & Cole, 2017). Although alternate populations could impact the expression of sex indicators, a good sex indicator should capture more variance in sex and be less impacted by population-level differences, especially if the expression is due to a selective force present in all populations (e.g., reproduction). A good example of this has been the continued success of the pelvis, and specifically the associated Phenice (1969) traits, in validation studies around the world and through time (Garcia de Leon & Toon, 2014; Klales, 2016; Lesciotto & Doershuk, 2018; Oikonomopoulou, Valakos, & Nikita, 2017). It is not to say that population variation is not present, as one can tell from recalibration studies (Gómez-Valdés et al., 2017; Kenyhercz et al., 2017; Klales & Cole, 2017), but rarely do the classification accuracies associated with the original publication change as much as they do for the subadult studies. As indicators decrease in predictive power, we would see a larger influence from population variation and greater variation among populations.

If a subadult sex indicator is associated with body size, the above argument may not be valid. Children from different populations exhibit considerable differences in body size and shape and in rate of growth, which gives rise to the population-level body size variation appreciated in adults (Eveleth, 1978). Because growth rates have a direct impact on SSD, it is true that some population disparities may be due to differential gene expression or environmental conditions that elicit a specific growth pattern. For example, SES explains almost none of the variance in body length at birth, but by 36 months, it can account for approximately 40% of body length (Eveleth, 1978). As age increases, the impact of environmental insults also increases. Differences have been identified in the degree of SSD between low SES and high SES groups, such that low SES individuals do not express statistically significant SSD until later in age compared to high SES individuals (Nikitovic & Bogin, 2014). The most likely reason for these findings is a slower growth rate in males from the low SES group experiencing greater environmental insult. The findings support the female buffering hypothesis that argues that poor environmental conditions decrease overall sex differences (in adults) by affecting males more than females.

Size differences in association with environmental influences have even been documented in skeletal elements that are considered to be under strong genetic control. Tooth size has been shown to have significant differences based on being preterm or full-term gestation. Individuals considered very low birthweight have teeth present with the smallest dimensions, while those considered normal birthweight express the largest teeth (Garn, Osborne, & McCabe, 1979; Seow & Wan, 2000).

Case study: Comparing long bone dimensions from South Africa and the United States

To further address the comparison of multiple populations, we provide an example of diaphyseal dimension data available from two countries with individuals of the same age. Sympercents (sp), or symmetric percentage differences, were calculated to quantify sexually dimorphic differences per measurement and age (Eq. 1). The analysis was conducted by age to account for the potentially different magnitudes between males and females that could become apparent throughout the growth period. The sp equation removes size by removing the unit of measure and allows for males and females to be symmetrically larger or smaller than one another because there is no denominator (Cole, 2000; Wells, 2012). Ultimately, sp offers an easier interpretation of dimorphic differences than the normal approach to demonstrating dimorphic differences, which is by percent differences (Cole, 2000; Stull, 2013). For example, when conducting percent differences the classic way, a group of females may be 14% shorter than a group of males, but the males are 16% taller than the same group of females. When presented via sp, males are 15.6% larger than females, and females are 15.6% smaller than males.

As one can see in Fig. 1, similar expressions of sexual dimorphism were observed in both a US sample ($n=784$; $F=315$, $M=466$) and a South African (SA) sample ($n=1310$; $F=506$, $M=804$). The level of sexual dimorphism differs between the two samples as they near 9 years of age.

$$(100(\log_e x_2)) - (100(\log_e x_1)) = \text{sp} \tag{1}$$

The similar growth trajectory of tibia maximum length for both SA and US individuals is apparent for populations and sexes until approximately 6 years (Fig. 2).

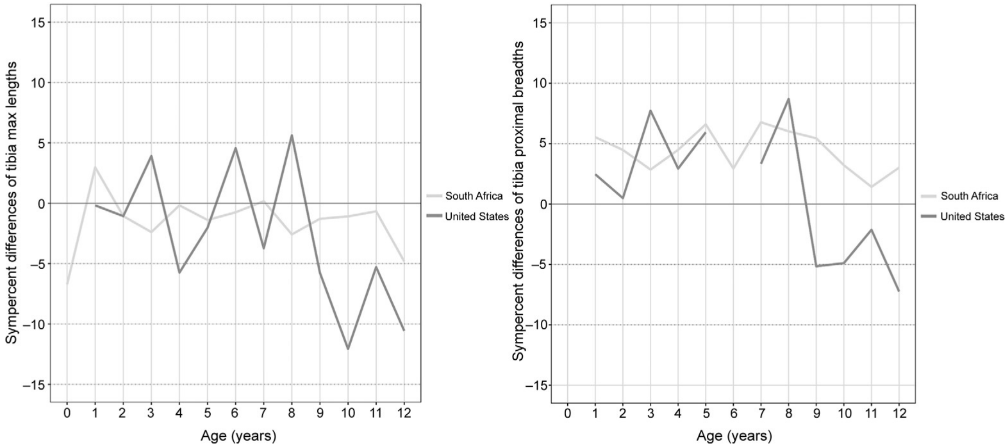


Fig. 1 Sympercent differences of the maximum tibia length (left) and the proximal tibia breadth (right) between females and males from South Africa and the United States.

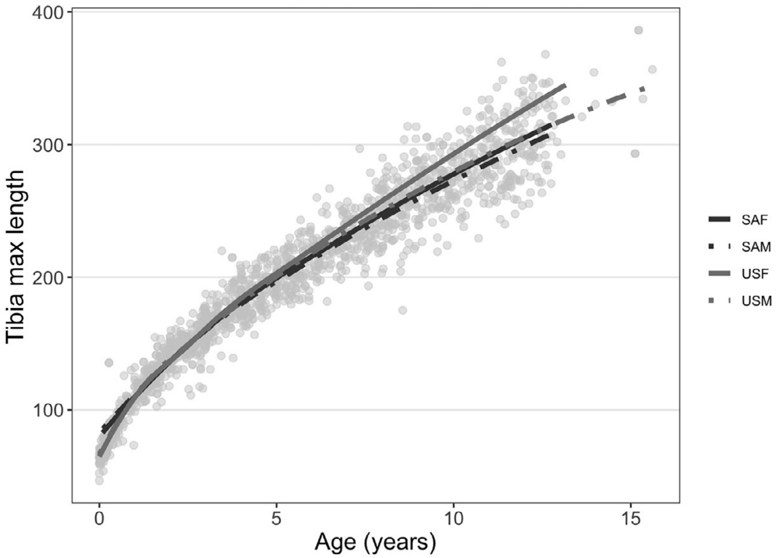


Fig. 2 Scatterplot with loess lines illustrating the gradual increase in average population and sex differences for both US and SA samples. Linetype differentiates the sexes, and color differentiates the country. SAF = South African females, SAM = South African males, USF = US females, USM = US males.

Around 6 years of age, the populations begin differentiating. As age increases, SA males and females present with, on average, shorter maximum tibia lengths per age than US males and females. US males and SA females track similarly through age progression. With increased age comes a large difference between US males and females; while a difference exists between SA males and females, it is smaller than the observed difference in the US sample. The findings here mimic those from Nikitovic and Bogin (2014) and elucidate a gradual increase in body size differences as well as the reduced expression of sexually dimorphic traits in the SA population, which can also be observed in Fig. 1. While only measurements of the tibia are illustrated, the trend was consistent across all long bones.

Sample sizes and predictive power of indicators

Although there are some inconsistencies, most of the methods discussed are strongly biased toward male classification. Essentially, males present with very high classification accuracies; in contrast, females present with very low classification accuracies. The inconsistency highlights some methodological considerations, including trait variation and what makes for the trait's capacity to discriminate between the sexes. One of the reasons that many of the methods present with such strong sex bias has to do with the shared expression of a trait. The greater the predictive power a trait has, the less the shared expression between males and females. If there is much overlap between the two sexes in a trait due to low predictive power, then a greater number of individuals will be assigned to the sex with a larger sample size and inherently will present with a higher classification accuracy (Hanifah, Wijayanto, & Kurnia, 2015). Beyond the biases introduced as a result of low predictive power, the abovementioned scenario also introduces the impediments that imbalanced classes can also have on a trait. Many machine learning statistics are also susceptible to imbalanced classes, or the discrepancy in sample sizes of outcome variables. This tendency to overclassify into the majority class is especially problematic in decision trees and logistic regression analyses. Researchers should look into resampling (e.g., downsampling, upsampling, SMOTE) methods to mitigate the discrepancy when even numbers of males and females cannot be sampled.

The additional sampling issue that one should consider is whether the distribution and size are accurate reflections of the larger population that one wants to estimate from. Data from all populations with vitality statistics demonstrate a female advantage in survival and higher death rates for males at virtually every single age (Oksuzyan, Juel, Vaupel, & Christensen, 2008; Wisser & Vaupel, 2014). As age increases, the mortality rate for males also increases. The differential mortality rate is exacerbated in a forensic population; the forensic sample has a much higher ratio of males to females compared to the larger population. Therefore, one must decide on the desired method they would like to develop. This discussion is especially important as forensic anthropologists are faced with more natural and mass disasters, and the ratio between males and females may be more balanced than the forensic sample.

When thinking of sample sizes and application, it is also inherently important to think about the indicators that are being utilized. If a method is using a sample distribution as the prior, the indicator will determine if it will result in a big or small impact on the final estimation. Essentially, imbalanced classes (and therefore an imbalanced prior) will have a large impact on a weak indicator, but only a small impact on a strong indicator. Similarly, the more data you have, the less reliance will be placed on the prior.

Incorporating multiple anatomical areas

Although biological and forensic anthropologists have been progressing toward the development of multiindicator (i.e., long bones and dentition) and multivariable (i.e., maxillary first molar and second molar) models, this has been largely limited to subadult and adult age estimation (e.g., Boldsen, Milner, Konigsberg, & Wood, 2002; Stull, L'Abbé, & Ousley, 2014). It is generally recognized that the incorporation of more indicators usually yields a more accurate estimate, which is especially the case when there is less predictive power in the indicators. Within subadult sex estimation research, only diaphyseal dimensions (Stull et al., 2017) and dentition (Viciano et al., 2013) have been used in a multivariable model. Methods that incorporated more than one variable consistently yielded accuracies that were higher than single-variable methods. For example, when both deciduous and adult dentition were used, higher accuracies were achieved (Viciano et al., 2013). These results support that a multivariable model would be superior to a single-indicator model, especially if the indicators themselves are not the strongest. Furthermore, if numerous anatomical areas express differences between the sexes but the general predictive power is weak, then a multi-indicator model will likely outperform a single-indicator model. There has been no research, to the authors' knowledge, that has looked to incorporate multiple indicators in sex estimation models.

Software for performing subadult sex estimation: KidStats

Stull, L'Abbe, and Ousley (2017) introduced KidStats, a graphical user interface (GUI), to facilitate application of flexible discriminant analysis when estimating subadult sex using upwards of 18 dimensions collected from long bones. The GUI allows the user to input available measurements and provides the classification results with the option of bootstrapping (Stull et al., 2017). The current version of KidStats is being updated to increase the number of reference samples and to increase the number and type of predictor variables, including both an ontogenetic approach using adult sex indicators as well as the continued option of using diaphyseal dimensions. The increase in reference samples now allows the users to run an estimate using a specific population, South Africa or United States, or a pooled sample with the data combined. All data was collected from full-body computed tomography images generated at the University of New Mexico Health Sciences Center, Office of the Medical Investigator and the Office of the Chief Medical Examiner, State of Maryland.

The preliminary results on the US sample, using models developed with 18 variables, downsampling for a balanced sex distribution, and using subsets for training ($n = 270$) and testing ($n = 88$), consistently demonstrates a 75% accuracy with no systematic misclassification patterns by age (Stull, Garvin, & Klales, 2020). While the majority of the South African sample does not contain all measurements, a subset of variables was used to demonstrate the performance when the reference samples were pooled. Using a smaller number of variables ($n = 8$), but still retaining the downsampling for a balanced sex distribution, and a training ($n = 390$) and testing ($n = 128$) subset, the models achieved on slightly lower classification accuracy (73%).

While the long bones are available for utilization, the US sample also offers an ontogenetic approach that can be applied in tandem with the long bones or individually. A portion of the US sample ($n = 301$) aged between 8 and 21 years was used in a pilot study evaluating the developmental trajectory of adult sex indicators including the Klales, Ousley, and Vollner (2012) and Walker (2008) traits and associated methodologies. The morphological traits of the innominate could successfully estimate sex in individuals as young as 13 years (86% and 93% classification accuracy for females and males, respectively), which was comparable to previous findings by Klales and Burns (2017) and Cole and Stull (2020a). However, the indicators observed on subadults do not achieve classification accuracies comparable to adults until approximately 15 years of age. The cranium tends to reach adult expressions later than the pelvis; trait score frequencies for subadult males were similar to those observed in adult males by 17 years (Cole & Stull, 2020b). Both anatomical areas have trait expressions prior to skeletal maturation and therefore subadult sex estimation can be performed using adult sex indicators. After these encouraging preliminary findings, additional data is being collected to bolster sample size and ensure the variation in the US is properly captured.

The improvements in methodology (e.g., downsampling and training and testing subsets) as well as additional reference samples are being made to KidStats to facilitate subadult sex estimation and to address some of the limitations of previous research mentioned above.

Conclusion

The most obvious conclusion when delving into the subadult literature is the lack of theoretical reasoning for much of the conducted research. When the data collected does not satisfy measurement theory criteria, the results cannot be interpreted, and on a larger scale, a critical evaluation of the potential of subadult sex estimation cannot be made. In the research that has been conducted, there is consistent lower performance of validation studies compared to original studies. While validation studies tend to have less success, there are substantial differences in the subadult sex literature, which only perpetuates the belief that subadult sex estimation is not possible in biological and forensic anthropology. Instead, the authors would argue that the capacity for estimating sex from the subadult skeleton has been underestimated. While the current state of the field may not offer robust methods that practitioners are confident in applying, the authors believe that the slow progress is due to larger obstacles (*i.e.*, lack of samples and improper methods) that have impeded the growth and popularity of this topic. As virtual collections become more popular, more aggressive analytics can be employed, and the ability to develop and test hypotheses will be substantially improved, as well as our understanding of the biomechanical, hormonal, and/or reproductive functions informing the expression of sex differences. The means that to ask and answer research questions will propel us closer to being able to identify sex indicators and differentiate what is theoretically causing sex differences from what is actually causing sex differences. Researchers will be able to test sex indicators throughout ontogeny and from widely varying geographic and economic populations, which will inform anthropologists on the expression of sex indicators that are under strong selective pressures and genetic control or susceptible to adverse environmental conditions. The authors believe that, in regard to subadult sex estimation, the absence of evidence is not the same as the evidence of absence, and we are excited for future developments that are yet to come. There is no doubt that improvements in subadult sex estimation will have tangible effects to the entire field of biological anthropology.

Acknowledgments

A special thanks to Dr. Louise Corron and Chris Wolfe for their critiques and helpful comments, and to Dr. Alexandra R. Klales for inviting us to participate in this greatly needed volume. Research reported in this chapter was supported by the National Institute of Justice under award numbers 2015-DN-BX-K409 and 2015-DN-BX-K009. Opinions or points of view expressed represent a consensus of the authors and do not necessarily represent the official position of the US Department of Justice.

References

Abbassi, V. (1998). Growth and normal puberty. *Pediatrics*, *102*, 507–511.

Acsadi, G., & Nemeskeri, J. (1970). *History of human life span and mortality*. Budapest: Akademiai Kiado.

Aksglaede, L., Juul, A., Leffers, H., Skakkebæk, N. E., & Andersson, A. -M. (2006). The sensitivity of the child to sex steroids: Possible impact of exogenous estrogens. *Human Reproduction Update*, *12*, 341–349.

Alexander, G. M., & Wilcox, T. 2012. Sex differences in early infancy. *Child Development Perspectives*, *6*(4), 400–406.

Andras, N. L., & Stock, M. K. (2018). *Subadult sex estimation using multi-slice computed tomography scans of the ilium*. Available from: http://meeting.physanth.org/program/2018/session24/andras-2018-subadult-sex-estimation-using-multi-slice-computed-tomography-scans-of-the-ilium.html.

Arfai, K., Pitukcheewanont, P. D., Goran, M. I., Tavare, C. J., Heller, L., & Gilsanz, V. (2002). Bone, muscle, and fat: Sex-related differences in prepubertal children. *Radiology*, *224*, 338–344.

Badyaev, A. V. (2002). Growing apart: An ontogenetic perspective on the evolution of sexual size dimorphism. *Trends in Ecology and Evolution*, *17*, 369–378.

Bernstein, R. M. (2010). The big and small of it: How body size evolves. *American Journal of Physical Anthropology*, *143*(Suppl. 51), 46–62.

Bilfeld, M. F., Dedouit, F., Sans, N., Rousseau, H., Rougé, D., & Telmon, N. (2013). Ontogeny of size and shape sexual dimorphism in the ilium: A multislice computed tomography study by geometric morphometry. *Journal of Forensic Sciences*, *58*, 303–310.

Bilfeld, M. F., Dedouit, F., Sans, N., Rousseau, H., Rougé, D., & Telmon, N. (2015). Ontogeny of size and shape sexual dimorphism in the pubis: A multislice computed tomography study by geometric morphometry. *Journal of Forensic Sciences*, *60*, 1121–1128.

Black, T. (1978). Sexual dimorphism in the tooth-crown diameters of the deciduous teeth. *American Journal of Physical Anthropology*, *48*, 77–82.

Bogin, B. (1999). *Patterns of human growth*. Cambridge: Cambridge University Press.

Boldsen, J., Milner, G., Konigsberg, L., & Wood, J. (2002). Transition analysis: A new method for estimating age from skeletons. In R. Hoppa & J. Vaupel (Eds.), *Paleodemography: Age distributions from skeletal samples* (pp. 73–106). Cambridge: Cambridge University Press.

Burgoyne, P. S., Thornhill, A. R., Boudrean, S. K., Darling, S. M., Bishop, C. E., Evans, E. P., et al. (1995). The genetic basis of XX-XY differences present before gonadal sex differentiation in the mouse [and discussion]. *Philosophical Transactions: Biological Sciences*, *350*, 253–261.

Cameron, N., Tanner, J. M., & Whitehouse, R. H. (1982). A longitudinal analysis of the growth of limb segments in adolescence. *Annals of Human Biology*, *9*, 211–220.

Cardoso, H. F. V. (2008). Sample-specific (universal) metric approaches for determining the sex of immature human skeletal remains using permanent tooth dimensions. *Journal of Archaeological Science*, *35*, 158–168.

Cardoso, H. F. V. (2010). Testing discriminant functions for sex determination from deciduous teeth. *Journal of Forensic Sciences*, *55*, 1557–1560.

Cardoso, H. F. V., & Saunders, S. R. (2008). Two arch criteria of the ilium for sex determination of immature skeletal remains: A test of their accuracy and an assessment of intra- and inter-observer error. *Forensic Science International*, *178*, 24–29.

Chasiotis, A., Scheffer, D., Restemeier, R., & Keller, H. (1998). Intergenerational context discontinuity affects the onset of puberty. *Human Nature: An Interdisciplinary Biosocial Perspective*, *9*, 321.

Choi, S., & Trotter, M. (1970). Statistical study of the multivariate structure and race-sex differences of American White and Negro fetal skeletons. *American Journal of Physical Anthropology*, *33*, 307–312.

Chou, C. -C., Iwasa, Y., & Nakazawa, T. (2016). Incorporating an ontogenetic perspective into evolutionary theory of sexual size dimorphism. *Evolution*, *70*, 369–384.

Chumlea, W., Schubert, C., & Roche, A. (2003). Age at menarche and racial comparisons in US girls. *Pediatrics*, *111*, 110–113.

Cole, T. J. (2000). Sympercents: Symmetric percentage differences on the 100 loge scale simplify the presentation of log transformed data. *Statistics in Medicine*, *19*, 3109–3125.

Cole, T. J. (2003). The secular trend in human physical growth: A biological view. *Economics and Human Biology*, *1*, 161–168.

Cole, S., & Stull, K. (2020a). Application and evaluation of adult morphological sex traits using the subadult innominate. *Presented at the American Academy of Forensic Sciences, Anaheim, CA*.

Cole, S., & Stull, K. (2020b). Quantifying the emergence of sexually dimorphic traits using Walker (2008) and evaluating method performance in subadult sex estimation. *Presented at the American Association of Physical Anthropologists, Los Angeles, CA*.

Cox, R. M., Stenquist, D. S., & Calsbeek, R. (2009). Testosterone, growth and the evolution of sexual size dimorphism. *Journal of Evolutionary Biology*, *22*, 1586–1598.

Danker-Hopfe, H. (1986). Menarcheal age in Europe. *American Journal of Physical Anthropology*, *29*, 81–112.

De Vito, C., & Saunders, S. (1990). A discriminant function analysis of deciduous teeth to determine sex. *Journal of Forensic Sciences*, *35*, 845–848.

DiGangi, E., & Moore, M. (Eds.), (2012). *Research methods in human skeletal biology* (1st ed.): Academic Press. Available from: https://www.elsevier.com/books/research-methods-in-human-skeletal-biology/digangi/978-0-12-385189-5.

Estévez Campo, E. J., López-Lázaro, S., López-Morago Rodríguez, C., Alemán Aguilera, I., & Botella López, M. C. (2018). Specific-age group sex estimation of infants through geometric morphometrics analysis of pubis and ischium. *Forensic Science International*, *286*, 185–192.

Euling, S. Y., Selevan, S. G., Pescovitz, O. H., & Skakkebaek, N. E. (2008). Role of environmental factors in the timing of puberty. *Pediatrics*, *121*, S167–S171.

Eveleth, P. (1978). Population differences in growth. environmental and genetic factors. In F. Falkner & J. Tanner (Eds.), *Vol. 1. Human growth* (pp. 221–239). New York: Plenum.

Freedman, D. S., Khan, L. K., Serdula, M. K., Ogden, C. L., & Dietz, W. H. (2006). Racial and ethnic differences in secular trends for childhood BMI, weight, and height. *Obesity*, *14*, 301–308.

Garcia de Leon, J., & Toon, C. (2014). *A comparison of the Klales et al. (2012) and Phenice (1969) methods of sex estimation on a modern Colombian sample*. In: American Academy of Forensic Science, 66th annual scientific meeting, Seattle, WA.

Garn, S. M., Cole, P. E., Wainwright, R. L., & Guire, K. E. (1977). Sex discriminatory effectiveness using combinations of permanent teeth. *Journal of Dental Research*, *56*, 697.

Garn, S. M., Osborne, R. H., & McCabe, K. D. (1979). The effect of prenatal factors on crown dimensions. *American Journal of Physical Anthropology*, *51*, 665–678.

Garvin, H., Severa, K., Ternent, E., Stock, M., Cirillo, L., & Stull, K. (2019). Assessing the potential of ilium outlines, greater sciatic notch metrics, and indices of pubic/ischial length for subadult sex estimation. In *Annual meeting of the American Association of Physical Anthropologists. Cleveland, OH*.

Golub, M. S., Collman, G. W., Foster, P. M. D., Kimmel, C. A., Meyts, E. R. -D., Reiter, E. O., et al. (2008). Public health implications of altered puberty timing. *Pediatrics*, *121*, S218–S230.

Gómez-Valdés, J. A., Menéndez Garmendia, A., García-Barzola, L., Sánchez-Mejorada, G., Karam, C., Baraybar, J. P., et al. (2017). Recalibration of the Klales et al. (2012) method of sexing the human innominate for Mexican populations: GÓmez-valdÉs et al. *American Journal of Physical Anthropology*, *162*, 600–604.

Grumbach, M. M. (2002). The neuroendocrinology of human puberty revisited. *Hormone Research*, *57*, 2–14.

Hanifah, F. S., Wijayanto, H., & Kurnia, A. (2015). SMOTE bagging algorithm for imbalanced dataset in logistic regression analysis (case: Credit of bank X). *Applied Mathematical Sciences*, *9*, 6857–6865.

Harlan, W. R., Grillo, G. P., Cornoni-Huntley, J., & Leaverton, P. E. (1979). Secondary sex characteristics of boys 12 to 17 years of age: The U.S. Health Examination Survey. *The Journal of Pediatrics*, *95*, 287–289.

Harlan, W. R., Harlan, E. A., & Grillo, G. P. (1980). Secondary sex characteristics of girls 12 to 17 years of age: The U.S. Health Examination Survey. *The Journal of Pediatrics*, *96*, 1074.

Hassett, B. (2011). Technical note: Estimating sex using cervical canine odontometrics: A test using a known sex sample. *American Journal of Physical Anthropology*, *146*, 486–489.

Hauspie, R. C., Vercauteren, M., & Susanne, C. (1997). Secular changes in growth and maturation: An update. *Acta Paediatrica*, *86*, 20–27.

Herman-Giddens, M. E., Slora, E. J., Wasserman, R. C., Bourdony, C. J., Bhapkar, M. V., Koch, G. G., et al. (1997). Secondary sexual characteristics and menses in young girls seen in office practice: A study from the pediatric research in office settings network. *Pediatrics*, *99*, 505–512.

Herman-Giddens, M. E., Steffes, J., Harris, D., Slora, E., Hussey, M., Dowshen, S. A., et al. (2012). Secondary sexual characteristics in boys: Data from the pediatric research in office settings network. *Pediatrics*, *130*, e1058–e1068.

Houle, D., Pelabon, C., Wagner, G., & Hansen, T. (2011). Measurement and meaning in biology. *The Quarterly Review of Biology*, *86*, 3–34.

Hunt, D. R. (1990). Sex determination in the subadult ilia: An indirect test of Weaver's nonmetric sexing method. *Journal of Forensic Sciences*, *35*, 881–885.

Hutt, C. (1972a). Sex differences in human development. *Human Develop*, *15*, 153–170.

Hutt, C. (1972b). *Males and females*. Oxford, England: Penguin.

Irurita Olivares, J., & Alemán Aguilera, I. (2016). Validation of the sex estimation method elaborated by Schutkowski in the Granada Osteological Collection of identified infant and young children: Analysis of the controversy between the different ways of analyzing and interpreting the results. *International Journal of Legal Medicine-Heidelberg*, *130*, 1623–1632.

Kashimada, K., & Koopman, P. (2010). Sry: The master switch in mammalian sex determination. *Development*, *137*, 3921–3930.

Kenyhercz, M. W., Klales, A. R., Stull, K. E., McCormick, K. A., & Cole, S. J. (2017). Worldwide population variation in pelvic sexual dimorphism: A validation and recalibration of the Klales et al. method. *Forensic Science International*. *277*, 259.e1–259.e8.

Klales, A. R. (2016). Secular change in morphological pelvic traits used for sex estimation. *Journal of Forensic Sciences*. Available from: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1556-4029.13008.

Klales, A. R., & Burns, T. L. (2017). Adapting and applying the Phenice (1969) adult morphological sex estimation technique to subadults. *Journal of Forensic Sciences*, *62*, 747–752.

Klales, A. R., & Cole, S. J. (2017). Improving nonmetric sex classification for Hispanic individuals. *Journal of Forensic Sciences*, *62*, 975–980.

Klales, A. R., Ousley, S. D., & Vollner, J. M. (2012). A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *American Journal of Physical Anthropology*, *149*(1), 104–114.

Klepinger, L. L. (2001). Stature, maturation variation and secular trends in forensic anthropology. *Journal of Forensic Sciences*, *46*, 788–790.

Lanciotti, L., Cofini, M., Leonardi, A., Penta, L., & Esposito, S. (2018). Up-to-date review about minipuberty and overview on hypothalamic-pituitary-gonadal axis activation in fetal and neonatal life. *Frontiers in Endocrinology (Lausanne)*, *9*. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6070773/.

Largo, R., Walli, R., Duc, G., Fanconi, A., & Prader, A. (1980). Evaluation of perinatal growth. Presentation of combined intra- and extrauterine growth standards for weight, length and head circumference. *Helvetica Paediatrica Acta*, *35*, 419–436.

Lee, Y., & Styne, D. (2013). Influences on the onset and tempo of puberty in human beings and implications for adolescent psychological development. *Hormones and Behavior*, *64*, 250–261.

Leigh, S. R., & Shea, B. T. (1995). Ontogeny and the evolution of adult body size dimorphism in apes. *American Journal of Primatology*, *36*, 37–60.

Lesciotto, K. M., & Doershuk, L. J. (2018). Accuracy and reliability of the Klales et al. (2012) morphoscopic pelvic sexing method. *Journal of Forensic Sciences*, *63*, 214–220.

López-Lázaro, S., Alemán, I., Viciano, J., & Irurita, J. (2018). Sexual dimorphism of the first deciduous molar: A geometric morphometric approach. *Forensic Science International*, *290*, 94–102.

Loth, S. R. (1996). Sexual dimorphism in the human mandible: An evolutionary and developmental perspective. In Ph.D. thesis, University of Witwatersrand. Ann Arbor: University Microfilms, Inc.

Loth, S., & Henneberg, M. (2001). Sexually dimorphic mandibular morphology in the first few years of life. *American Journal of Physical Anthropology*, *115*, 179–186.

Lubchenco, L., Hansman, C., & Boyd, E. (1966). Intrauterine growth in length and head circumference as estimated from live births at gestational ages from 26 to 42 weeks. *Pediatrics*, *37*, 403–408.

Lubchenco, L., Hansman, C., Dressler, M., & Boyd, E. (1963). Intrauterine growth as estimated from liveborn birth-weight data at 24 to 42 weeks of gestation. *Pediatrics*, *32*, 793–800.

Lund, H., & Mörnstad, H. (1999). Gender determination by odontometrics in a Swedish population. *The Journal of Forensic Odonto-Stomatology*, *17*, 30–34.

Mann, D., & Fraser, H. (1996). The neonatal period: A critical interval in male primate development. *The Journal of Endocrinology*, *149*, 191–197.

Margetts, B., & Brown, T. (1978). Crown diameters of the deciduous teeth in Australian Aboriginals. *American Journal of Physical Anthropology*, *48*, 493–502.

Matthews, H. S., Penington, A. J., Hardiman, R., Fan, Y., Clement, J. G., Kilpatrick, N. M., et al. (2018). Modelling 3D craniofacial growth trajectories for population comparison and classification illustrated using sex-differences. *Scientific Reports*, *8*, 4771.

McClelland, K., Bowles, J., & Koopman, P. (2012). Male sex determination: Insights into molecular mechanisms. *Asian Journal of Andrology*, *14*, 164–171.

Mittler, D. M., & Sheridan, S. G. (1992). Sex determination in subadults using auricular surface morphology: A forensic science perspective. *Journal of Forensic Sciences*, *37*, 13293J.

Mittwoch, U. (1993). Blastocysts prepare for the race to be male. *Human Reproduction (Oxford, England)*, *8*, 1550.

Moerman, M. L. (1982). Growth of the birth canal in adolescent girls. *American Journal of Obstetrics and Gynecology*, *143*, 528–532.

Molleson, T., Cruse, K., & Mays, S. (1998). Some sexually dimorphic features of the human juvenile skull and their values in sex determination in immature remains. *Journal of Archaeological Science*, *25*, 719–728.

Moorrees, C., Fanning, E., & Hunt, E. (1963). Age variation of formation stages for ten permanent teeth. *Journal of Dental Research*, *42*, 1490–1502.

Nikitovic, D., & Bogin, B. (2014). Ontogeny of sexual size dimorphism and environmental quality in Guatemalan children. *American Journal of Human Biology*, *26*, 117–123.

Oikonomopoulou, E. -K., Valakos, E., & Nikita, E. (2017). Population-specificity of sexual dimorphism in cranial and pelvic traits: Evaluation of existing and proposal of new functions for sex assessment in a Greek assemblage. *International Journal of Legal Medicine*, *131*, 1731–1738.

Oksuzyan, A., Juel, K., Vaupel, J. W., & Christensen, K. (2008). Men: Good health and high mortality. Sex differences in health and aging. *Aging Clinical and Experimental Research*, *20*, 91–102.

Olsen, I., Groveman, S., Lawson, M., Clark, R., & Zemel, B. (2010). New intrauterine growth curves based on United States data. *Pediatrics*, *125*, 214–224.

Ounsted, C., & Taylor, D. C. (Eds.), (1972). *Gender differences: Their ontogeny and significance* (1st ed.). Edinburgh: Churchill Livingstone.

Ousley, S. (2013). *A radiographic database for estimating biological parameters in modern subadults. Final technical report, National Institute of Justice Award Number 2008-DN-BX-K152*. Available from: https://ncjrs.gov/pdffiles1/nij/grants/242697.pdfErie, PA: Mercyhurst University.

Phenice, T. (1969). A newly developed visual method of sexing the os pubis. *American Journal of Physical Anthropology*, *30*, 297–301.

Ray, P. F., Conaghan, J., Winston, R. M., & Handyside, A. H. (1995). Increased number of cells and metabolic activity in male human preimplantation embryos following in vitro fertilization. *Journal of Reproduction and Fertility*, *104*, 165.

Rhen, T. (2007). *Sex differences: Genetic, physiological, and ecological mechanisms*. Oxford University Press. Available from: http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199208784.001.0001/acprof-9780199208784-chapter-19.

Scheuer, L. (2002). A blind test of mandibular morphology for sexing mandibles in the first few years of life. *American Journal of Physical Anthropology*, *119*, 189–191.

Schiessl, H., Frost, H. M., & Jee, W. S. S. (1998). Estrogen and bone-muscle strength and mass relationships. *Bone*, *22*, 1–6.

Schutkowski, H. (1993). Sex determination of infant and juvenile skeletons: I. Morphognostic features. *American Journal of Physical Anthropology*, *90*, 199–206.

Scientific Working Group for Forensic Anthropology (SWGANTH) (2013). *Sex estimation*. Available from: https://www.nist.gov/system/files/documents/2018/03/13/swganth_sex_assessment.pdf.

Seow, W. K., & Wan, A. (2000). A controlled study of the morphometric changes in the primary dentition of pre-term, very-low-birthweight children. *Journal of Dental Research*, *79*, 63.

Shea, B. T. (1992). Developmental perspective on size change and allometry in evolution. *Evolutionary Anthropology: Issues, News, and Reviews*, *1*, 125–134.

Sørensen, K., Mouritsen, A., Aksglaede, L., Hagen, C. P., Mogensen, S. S., & Juul, A. (2012). Recent secular trends in pubertal timing: Implications for evaluation and diagnosis of precocious puberty. *Hormone Research in Paediatrics*, *77*, 137–145.

Stinson, S. (1985). Sex differences in environmental sensitivity during growth and development. *American Journal of Physical Anthropology*, *28*, 123–147.

Stock, M. K. (2018). A preliminary analysis of the age of full expression of sexually dimorphic cranial traits. *Journal of Forensic Sciences*, *63*, 1802–1808.

Stock, M. K., Stull, K. E., Garvin, H. M., & Klales, A. R. (2016). Development of modern human subadult age and sex estimation standards using multi-slice computed tomography images from medical examiner's offices. In *Vol. 9967. Proc. SPIE 9967, developments in X-ray tomography X* (pp. 99670E–99670E-14). https://doi.org/10.1117/12.2237180.

Stull, K. (2013). An osteometric evaluation of age and sex differences in the long bones of South African Children from the Western Cape. PhD Dissertation. South Africa: University of Pretoria.

Stull, K., Garvin, H., & Klales, A. (2020). *Final report: Development of modern subadult standards: Improved age and sex estimation in U.S. Forensic Practice* (Grant No. Final Report NIJ Award 2015-DN-BX-K409). Reno, NV: University of Nevada, Reno.

Stull, K. E., & Godde, K. (2013). Sex estimation of infants between birth and one year through discriminant analysis of the humerus and femur. *Journal of Forensic Sciences*, *58*(1), 13–20.

Stull, K., L'Abbe, E., & Ousley, S. (2017). Subadult sex estimation from diaphyseal dimensions. *American Journal of Physical Anthropology*, *163*(1), 64–74. https://doi.org/10.1002/ajpa.23185.

Stull, K. E., L'Abbé, E. N., & Ousley, S. D. (2014). Using multivariate adaptive regression splines to estimate subadult age from diaphyseal dimensions. *American Journal of Physical Anthropology 154*(3), 376–386.

Stulp, G., & Barrett, L. (2016). Evolutionary perspectives on human height variation: Evolution and human height. *Biological Reviews*, *91*, 206–234.

Styne, D., & Grumbach, M. (2011). Puberty: Ontogeny, neuroendocrinology, physiology, and disorders. In S. Melmed, K. Polonsky, P. Larsen, & H. Kronenberg (Eds.), *Williams textbook of endocrinology* (12th ed.). Philadelphia: Saunders Elsevier.

Sutter, R. (2003). Nonmetric subadult skeletal sexing Traits: l. A blind test of the accuracy of eight previously proposed methods using prehistoric known-sex mummies from Northern Chile. *Journal of Forensic Sciences*, *48*, 927–935.

Tanner, J. (1989). *Foetus into man: Physical growth from conception to maturity*. Cambridge: Howard University Press.

Tanner, J. M. (1992). Growth as a measure of the nutritional and hygienic status of a population. *Hormone Research*, *38*(Suppl. 1), 106.

Tanner, J. M., Hughes, P. C., & Whitehouse, R. H. (1981). Radiographically determined widths of bone muscle and fat in the upper arm and calf from age 3-18 years. *Annals of Human Biology*, *8*, 495–517.

Thomas, P., Peabody, J., Tunier, V., & Clark, R. (2000). A new look at intrauterine growth and the impact of race, altitude, and gender. *Pediatrics*, *106*.

Viciano, J., López-Lázaro, S., & Alemán, I. (2013). Sex estimation based on deciduous and permanent dentition in a contemporary Spanish population. *American Journal of Physical Anthropology*, *152*, 31–43.

Walker, P. L. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, *136*(1), 39–50.

Weaver, D. S. (1980). Sex differences in the Ilia of a known sex and age sample of fetal and infant Skeletons. *American Journal of Physical Anthropology*, *52*, 191–195.

Wells, J. C. K. (2007). Sexual dimorphism of body composition. *Best Practice and Research Clinical Endocrinology and Metabolism*, *21*, 415–430.

Wells, J. C. K. (2012). Sexual dimorphism in body composition across human populations: Associations with climate and proxies for short- and long-term energy supply. *American Journal of Human Biology*, *24*, 411–419.

Wilson, L., Cardoso, H., & Humphrey, L. (2011). On the reliability of a geometric morphometric approach to sex determination: A blind test of six criteria of the juvenile ilium. *Forensic Science International*, *206*, 35–42.

Wilson, L. A., MacLeod, N., & Humphrey, L. T. (2008). Morphometric criteria for sexing juvenile human skeletons using the ilium. *Journal of Forensic Sciences*, *53*, 269–278.

Wilson, L. A. B., & Humphrey, L. T. (2017). Voyaging into the third dimension: A perspective on virtual methods and their application to studies of juvenile sex estimation and the ontogeny of sexual dimorphism. *Forensic Science International—Amsterdam*, *278*, 32–46.

Wisser, O., & Vaupel, J. (2014). *The sex differential in mortality: A historical comparison of the adult-age pattern of the ratio and the difference*. Max Planck Institute for Demographic Research.

Zadzinska, E., Karasinska, M., Jedrychowska-Danska, K., Watala, C., & Witas, H. (2008). Sex diagnosis of subadult specimens from Medieval Polish archaeological sites: Metric analysis of deciduous dentition. *Homo-Journal of Comparative Human Biology*, *59*, 175–187.

CHAPTER 15

DSP: A probabilistic approach to sex estimation free from population specificity using innominate measurements

Frédéric Santos[a], Pierre Guyomarc'h[b], Eugénia Cunha[c], Jaroslav Brůžek[a,d]
[a]Université de Bordeaux—CNRS—MCC, UMR 5199 PACEA, Pessac, France
[b]Université Aix Marseille—CNRS—EFS, Marseille, France
[c]Centre for Functional Ecology, Laboratory of Forensic Anthropology, Life Sciences Department, University of Coimbra, Coimbra, Portugal
[d]Department of Anthropology and Human Genetics, Faculty of Science, Charles University, Prague, Czech Republic

Introduction

Biological anthropology globally recognizes that the most appropriate anatomical region for sexing unknown skeletonized human remains is the bony pelvis, despite its considerable fragility (e.g., Berg, 2017; Frayer & Wolpoff, 1985; İşcan & Steyn, 2013). The innominate bears the majority of pelvic sexual dimorphism. This is emphasized by all publications developing methods for the estimation of biological profile, and this is accepted in both forensic sciences and bioarchaeology. As a consequence, the most precise and reliable methods for sexing are based on the innominate.

The whole is more than just the sum of its parts

The theoretical framework for the study of pelvic sexual dimorphism and the design of the sex estimation tool *Diagnose Sexuelle Probabiliste* or DSP (Brůžek, Santos, Dutailly, Murail, & Cunha, 2017; Murail, Brůžek, Houët, & Cunha, 2005) was heavily influenced by the work of Novotný (1981, 1986). Inspired by the general systems theory (von Bertalanffy, 1968), Novotny split the pelvis into two large subsystems or segments: the sacro-iliac and ischio-pubic segments (or modules), the first consisting of the sacrum and ilium, and the second of the pubis and ischium. The innominate can be seen also as a relatively integrated unit consisting of two or three basic modules, namely a sacro-iliac module, an acetabular module, and an ischio-pubic module (Brůžek, 1991). The module that forms the ilium and sacrum reflects evolutionary adaptations to verticality and bipedalism; additionally, sex differences are the result of differential

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00015-8

adaptation to locomotion and reproduction between males and females, and the greater sciatic notch is the key structure of this module. The ischium/pubis proportions are the defining feature of the ischio-pubic module, and the acetabular module reflects biomechanical differences between the sexes, and the acetabular diameter is the representative feature of this module. Modularity helps to understand the evolvability and plasticity of organismal forms (Esteve-Altava, 2017a, 2017b; Klingenberg, 2008) and, consequently, sexual dimorphism and its variation. The results of Lewton (2012) and Grabowski, Polk, and Roseman (2011) strongly support the coexistence of the two modules, ischio-pubic and ilium, with a low level of integration.

"Rather than dividing a complex problem into its component parts, the systems perspective appreciates the holistic and composite characteristics of a problem and evaluates the problem with the use of computational and mathematical tools. The systems perspective is rooted in the assumption that the forest cannot be explained by studying the trees individually" (Ahn, Tewari, Poon, & Phillips, 2006, p. 709). Therefore, sex estimation methods should preferably contain variables covering the three modules. In addition, these variables should not be in strong correlation and should be clearly defined. It is only under such conditions that a sex estimation method can fully benefit from the sexual dimorphism of pelvic bone as a whole, and attain optimal reliability. Brůžek (1992) demonstrated this aspect by testing different discriminant functions containing the dimensions of different modules: those that did not meet the abovementioned hypotheses proved to be strongly population-specific. For these reasons, the selection of variables must receive a considerable attention.

Searching for reproducible and objective dimensions integrating sexual dimorphism

One of the most curious anomalies in the history of biological anthropology is the paucity of research on quantifying sex differences in the human pelvis in the early literature of the discipline. An element of explanation is that the 19th century's anthropologists were mostly preoccupied with problems of "race" (Hoyme, 1957). Sex differences in pelvic morphology were well known to anthropologists of the late 19th century (Dwight, 1878; Verneau, 1875). At the beginning of the 20th century, the typological approach and proposals of various indexes prevailed (Derry, 1923; Straus, 1927). The introduction of the ischio-pubic index (Schultz, 1930; Washburn, 1948, 1949), as a relationship between the length of pubis and the length of ischium, was a turning point in the historical evolution of sex estimation methods. The exact value of this index depends first and foremost on the possibility of precisely defining the lengths of the two bones, i.e., the position of the acetabular *point A* on which the three bones that form the innominate meet. However, the positioning of this landmark has a low accuracy since there is rarely a remaining clear trace of ossification of the zone in adulthood.

Pubis/ischium proportions measurements

The main problem in measuring the pubis length and ischium length is the determination of the acetabular point *A*, in which the ilium, ischium, and pubis meet (e.g., Adams & Byrd, 2002; Brůžek, 1984; Drew, 2013; Kim, Lee, Han, & Lee, 2018; Novotný, 1981). The difficulty to position this point *A* gave rise to a whole series of proposals (Fig. 1), the aim of which was to obtain a better approximation of its position (Bräuer, 1988; Gaillard, 1960, 1961; Genovés, 1959; Moeschler, 1964; Washburn, 1948, 1949).

Dimensions that start from the center of the acetabulum (Segebarth-Orban, 1980), and measure the so-called biomechanical pubis length and ischium length, are inappropriate for sex estimation. After Schultz (1930), this point *A* lies in all primates approximately at the intersection of the inner edge of facies lunata with a straight line

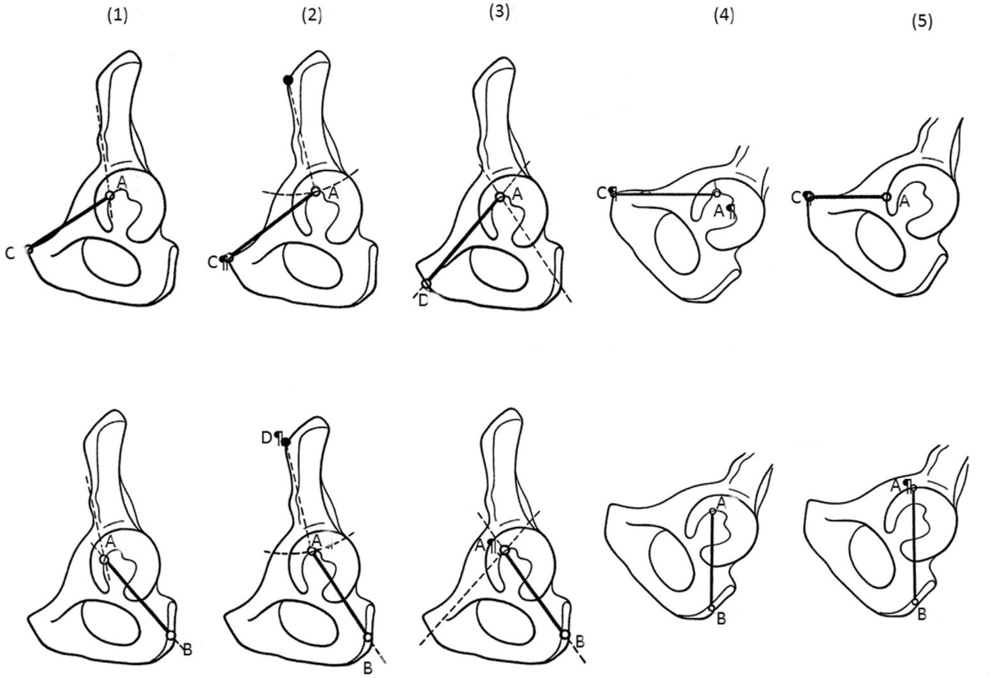


Fig. 1 Historical proposals for searching the most appropriate way of measuring the pubis and ischium lengths. Top line: pubis length, i.e., distance between symphysion (C) and acetabular point (A). Bottom line: ischium length, i.e., distance between ischial tuberosity and acetabular point (A). Five proposals are illustrated: 1. Schultz (1930); 2. Genovés (1959); 3. Gaillard (1961) and Seidler (1980); 4. Washburn (1948), Buikstra and Ubelaker (1994), and Byers (2002); 5. Thieme and Schull (1957), Novotný (1981), and Brůžek (1991). Four points are presented: A, acetabular point as defined by different authors (A1–A5); B, intersection between the axis of the superior ischial ramus and ischial tuberosity; C, symphysion, i.e., the upper end of pubis symphysis; D, intersection of the axis of ramus superior ossis pubis with the edge of symphysis.

prolonging the lower part of the acetabular border of the ilium downward. Another way is the definition of the acetabular point according to Genovés (1959), where point *A* is on the inner edge of the upper arm of facies lunata, which is closest to the spina iliaca anterior superior. Further modification was proposed by Gaillard (1961), which defines point *A* as the intersection of the axis of ramus superior ossis pubis and the axis of ramus superior ossis ischii in the acetabulum. Washburn and others (Buikstra & Ubelaker, 1994; Washburn, 1948, 1949) describe this point *A* in a different ways as an irregularity, a change in bone translucency, or a notch of the articular surface in the acetabulum. Point *D* is the same for all pubis length measurement techniques (symphysion) except the Gaillard (1961). Thieme and Schull (1957) suggested the most appropriate solution. They suggest not using it at all but instead measure the pubis length as a direct distance from the point symphysion (*C*) to the nearest point of the articular surface in acetabulum (*A*).

The dissatisfaction brought by these proposals first led to the use of another landmark, the acetabular center (Bräuer, 1988; Segebarth-Orban, 1980; Seidler, 1980); nevertheless, this point is also difficult to locate with reproducibility. During this methodological confusion, pubis length and ischium length were both used and rejected. “Problematic measurements such as pubis length are invalid due to the problem of locating a particular landmark (i.e., the junction of the pubis, ischium, and ilium in the acetabulum); these measurements should not be used in analyses” (Langley et al., 2018). No matter the reason, omitting the ischio-pubic dimensions means excluding the most important variables involving sexual dimorphism, thus being relevant in sex estimation.

Thieme and Schull (1957) resolved this dilemma by totally removing the acetabular point *A*. Instead, they proposed to measure the length of the pubis between the pubic symphysis and the closest landmark on the acetabular border. As far as the length of the ischium is concerned, Thieme and Schull (1957) proposed to measure this variable between the ischial tuberosity and the furthest landmark on the acetabular border. This suggestion has been used by a number of authors (e.g., Brůžek, 1984, 1991; Kimura, 1982; Novotný, 1981, 1986) and is also included in the DSP (Brůžek et al., 2017; Murail et al., 2005).

Measurements of the greater sciatic notch

The shape of the greater sciatic notch is another key structure used in sex estimation. Studies of the relationships between dimensions of the sciatic notch were confronted with the wide variability of anatomical structures that delimit it (Genovés, 1959; Hanna & Washburn, 1953; Jovanović & Živanović, 1965; Lazorthes & Lhès, 1939; Letterman, 1941; Singh & Potturi, 1978). Difficulties in measuring these dimensions include variety in shape of the ischial spine, which is often damaged (Jovanović & Živanović, 1965), and the tubercle of the muscle piriformis that is not always formed

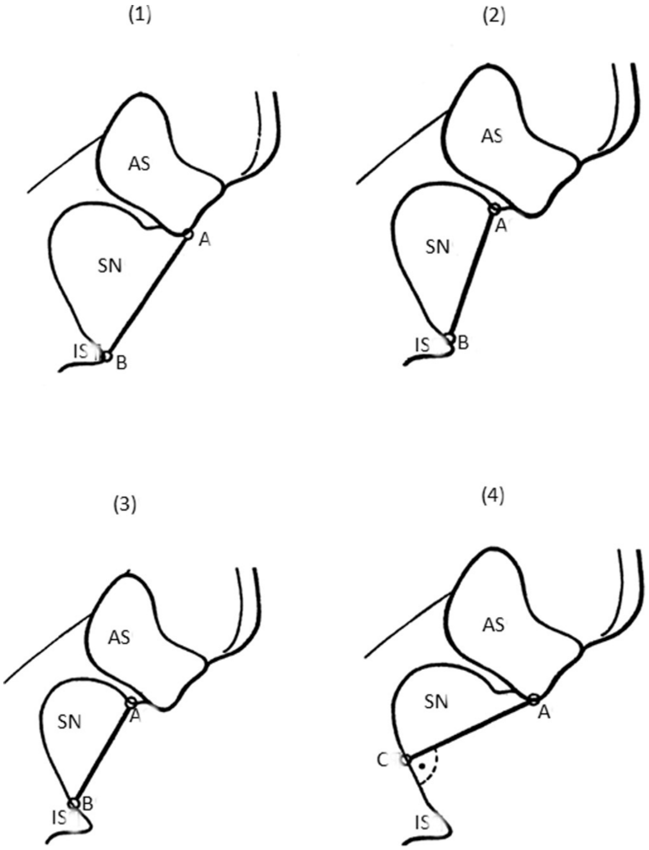


Fig. 2 Historical proposals for searching the most appropriate way of measuring the sciatic notch. AS, auricular surface; SN, sciatic notch; IS, ischial spine. Four proposals are illustrated: 1. Verneau (1875) and Kim et al. (2018): distance between the posterior inferior iliac spine (A) and the tip of ischial spine (B); 2. Lazorthes and Lhès (1939), Davivongs (1963), Singh and Potturi (1978): distance between the tip of tubercle of muscle pyriformis (A) and the tip of ischial spine (B); 3. Genovés (1959): distance between the tip of tubercle of muscle pyriformis (A) and the base of ischial spine (B); 4. Tin (1938): perpendicular distance between the posterior inferior iliac spine (A) and the inferior outline of greather sciatic notch (C). This variable is also called as the greater sciatic notch height (e.g., Bräuer, 1988; Brůžek et al., 2017).

(Lazorthes & Lhès, 1939); the most acceptable landmark for measurements is, therefore, the postero-inferior iliac spine (Fig. 2). Eventually, the solution proposed by Novotný (1981, 1986) proved temporarily satisfactory. He proposed to measure the width of the notch as the distance between the base of the ischial spine and the apex of the tubercle of muscle piriformis. In case of its absence, the tubercle is replaced by the point of contact between the upper edge of the notch and the end of auricular facies. This approach was followed by Brůžek (1992) and Brůžek, Murail, Houët, and Cleuvenot (1994).

However, during the process of variable selection for DSP (Brůžek et al., 2017; Murail et al., 2005), other measures of sciatic notch were preferred, like the greater sciatic notch height, defined by Tin (1938) and mentioned in Bräuer (1988).

Selection and final decision on the variables for DSP

As noted by İşcan and Steyn (2013), "through the last 80 years a number of authors have thus made significant contribution with regard to the development of several pelvic measurements and indices and the setting of standards for various populations." Among these measurements, the choice for DSP was mainly based on the results of dissertations from Novotný (1981) and Brůžek (1984), which sought to identify suitable dimensions to apply linear discriminant analyses (LDA) to the innominate. Novotny's analysis was based on a set of 36 variables. Brůžek (1991, 1992) studied Novotny's recommended dimensions in addition to some variables from the literature: a total of 32 variables were analyzed in two samples of innominates of known sex. Seventeen innominate measurements were further selected according to previous publications (Brůžek et al., 1994). From this pool of variables, the methodological approach to select optimal ones for DSP included three steps:

1. An investigation toward a common sexual dimorphism pattern among modern human populations, using LDA. Firstly, the definition of an overall European model was based on three European samples (London, Paris, and Coimbra); secondly, a test of the European model was performed on two independent Euro-American samples (Cleveland and Washington); thirdly, a new model including European and North American samples was built; finally, a test of the previous multiregional model was performed on African, Asian, and European (Vilnius) samples.
2. From the results, a selection of a subset of variables was made according to their discriminant power and their preservation rate.
3. The DSP tool for probabilistic sex estimation was developed using the pooled worldwide sample based on a subset of 10 optimal variables (Murail et al., 2005).

The aim of DSP is to propose a method for probabilistic sex estimation based on a broad dataset of innominate dimensions corresponding to the variability of the human species in time and space. The tool accepts the combination of a minimum of four dimensions out of the ten variables available. This holds value compared to LDA and other classification techniques that require a fixed number of dimensions, and that do not meet the practical requirement because of "the tyranny of taphonomy." Indeed, allowing for some flexibility in the use of variables minimizes the problem of bone preservation (Waldron, 1987). A good example of this issue is the funeral assemblage of *L'Isle Jourdan-La Gravette*, France, 6–12th century (Barthélémy, 1999). The whole assemblage included a total number of 819 skeletons, 527 of them being adult subjects, and innominates (or fragments) were preserved in only 343 cases. The estimation of sex using published and reliable discriminant functions was possible for only 74 individuals out of the 819. The rest of the chapter presents DSP and the new cross-platform software DSP2, available since 2017.

Material and methods

Samples

Calibration sample

The reference sample used for calibration of the model includes 2040 innominates from different geographic areas. This multiethnic and multipopulation sample was brought together to reflect the worldwide variation of pelvic morphology, looking for a common pattern of sexual dimorphism across all populations of modern humans. It constitutes the learning dataset for LDA models implemented in DSP2 software, which allow getting an accurate sex estimate of an unknown adult innominate. The details of data recorded from the 12 population samples can be found in Table 1. All individuals are adult subjects of known sex, and were measured between 1986 and 2002 (Murail et al., 2005).

Validation sample

A second sample is used for the validation of reliability of DSP2 on data independent from the calibration (reference) sample. It is composed of two series of adult innominates of known sex, extracted from two collections. The first series is composed of 120 innominates from the Maxwell Museum Documented Collection, University of New Mexico, Albuquerque, USA (Komar & Grivas, 2008). They belong to 61 identified adult skeletons of

Table 1 Number of innominates per subsample in the calibration dataset and target dataset (the number on individuals is indicated within parentheses).

	Subsample	**Females (*n*)**	**Males (*n*)**
Calibration sample	1	62 (31)	98 (49)
	2	130 (65)	102 (51)
	3	31 (31)	31 (31)
	4	112 (57)	108 (54)
	5	153 (78)	153 (79)
	6	58 (29)	52 (27)
	7	56 (28)	56 (29)
	8	56 (28)	56 (28)
	9	57 (29)	56 (28)
	10	102 (51)	97 (49)
	11	110 (55)	106 (53)
	12	96 (48)	102 (53)
	TOTAL	1023 (530)	1017 (531)
Validation sample	13	59 (30)	61 (31)
	14	250 (unknown)	253 (unknown)

Short codes for the collections: 1. Olivier Collection (Paris, France), 2. Tamagnini collection (Coimbra, Portugal), 3. Spitalfields (London, England), 4. Garmus collection (Vilnius, Lithuania), 5. Dart collection, Zulu (Johannesburg, South Africa), 6. Dart collection, Soto (Johannesburg, South Africa), 7. Dart collection, Afrikaner (Johannesburg, South Africa), 8. Hamann-Todd collection, Euroamerican (Cleveland, USA), 9. Hamann-Todd collection, Afroamerican (Cleveland, USA), 10. Terry collection, Euroamerican (Washington, DC, USA), 11. Terry collection, Afroamerican (Washington, DC, USA), 12. Chang-Mai collection, Thai (Chang-Mai), 13. Maxwell collection (Albuquerque, USA), 14. Simon collection (Geneva, Switzerland).

both sexes of American residents deceased at the end of the 20th century. The second series consists of 503 innominates from 503 identified individuals of both sexes from the Simon collection housed at the Department of Anthropology, University of Geneva, Switzerland. The individuals from this collection died between 1900 and 1969 and were buried in 27 cemeteries of the *Canton de Vaud* on the northern side of Lake Geneva (Henderson, Mariotti, Pany-Kucera, Villotte, & Wilczak, 2013; Perréard-Lopreno, 2007).

Measurements

DSP considers 10 measurements: eight of them are distributed on the three modules of the innominate, and two of them are general dimensions. Most of these dimensions are contained in the latest edition of Martin's manual (Bräuer, 1988), in which case their previous denomination is recalled below. All 10 measurements are regrouped by modules:

Innominate as a whole

DCOX (M1): innominate or coxal length (Bräuer, 1988) (Fig. 3).

SCOX (M12): Iliac or coxal breadth (Bräuer, 1988) (Fig. 4).

Ischio-pubic module

PUM (M14): acetabulo-symphyseal pubic length (Bräuer, 1988) (Fig. 5).

SPU: cotylo-pubic width (Gaillard, 1960) (Fig. 6).

ISMM: ischium postacetabular length (Schulter-Ellis, Schmidt, Hayek, & Craig, 1983) (Fig. 7).

Acetabular module

VEAC (M22): vertical acetabular diameter (Bräuer, 1988) (Fig. 8).

SIS (M14.1): cotylo-sciatic breadth (Bräuer, 1988) (Fig. 9).

Fig. 3 Coxal length.

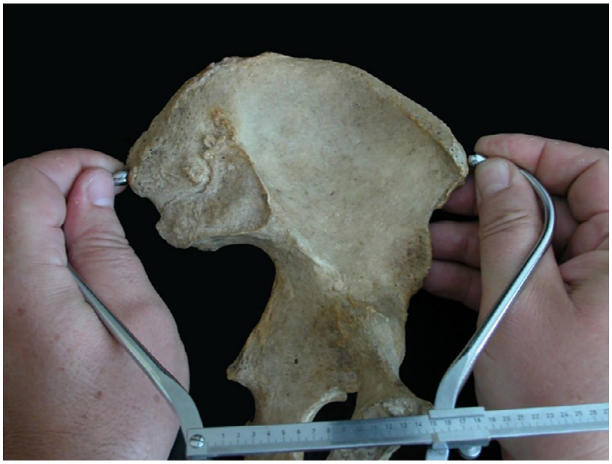


Fig. 4 Coxal breadth.

Fig. 5 Acetabulo-symphyseal pubic length.

Fig. 6 Cotylo-pubic width.

Fig. 7 Ischium postacetabular length.

Fig. 8 Vertical acetabular diameter.

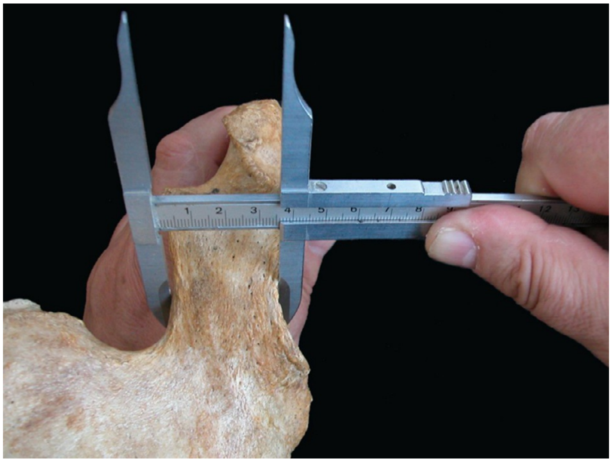


Fig. 9 Cotylo-sciatic breadth.

Sacro-iliac module

IIMT (M15.1): greater sciatic notch height (Bräuer, 1988) (Fig. 10).

SS: spino-sciatic length (Gaillard, 1960) (Fig. 11).

SA: spino-auricular length (Gaillard, 1960) (Fig. 12).

All these variables can be measured with the external jaws of a sliding caliper, except IIMT that rather requires the internal jaws of the caliper (or a friction divider). For the calibration (reference) and validation samples, all variables were measured by J. Brůžek. Intra- and interobserver errors were evaluated using the mean absolute difference (Utermohle & Zegura, 1982) and the technical error of measurement (Cameron, 1986) and showed an acceptable level of uncertainty (Brůžek et al., 1994; Vacca & Di Vella, 2012).

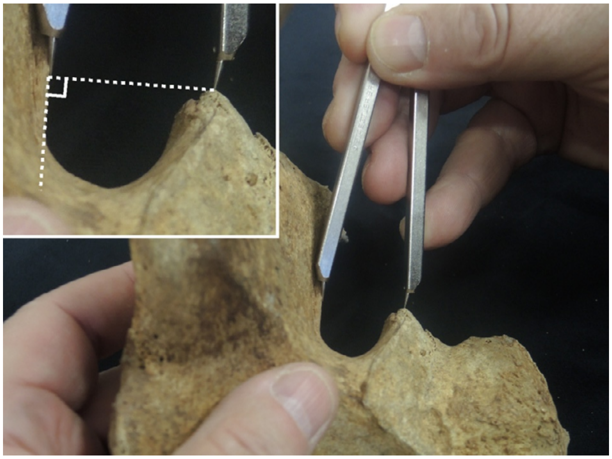


Fig. 10 Greater sciatic notch height.

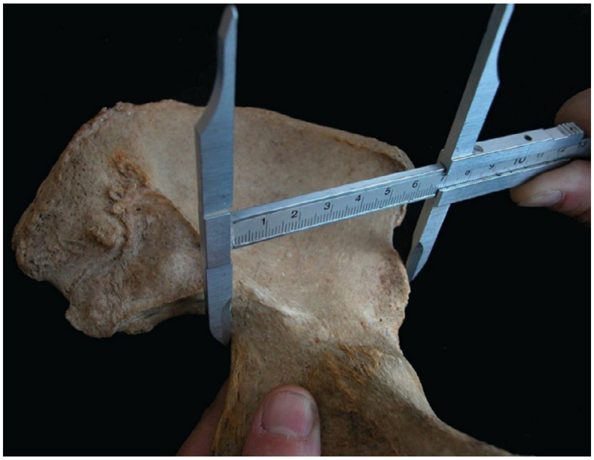


Fig. 11 Spino-sciatic length.

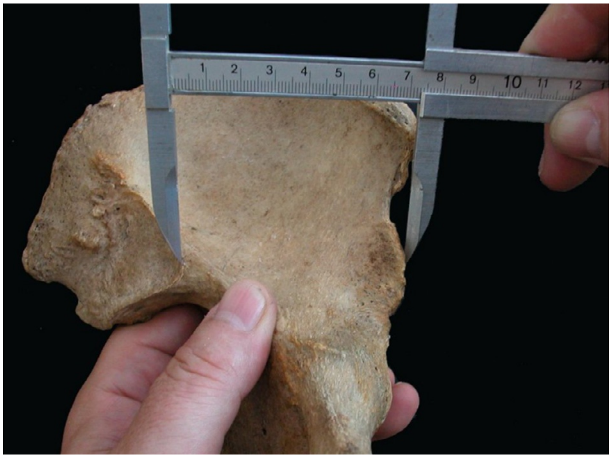


Fig. 12 Spino-auricular length.

Almost all measurements show a significant difference in means between males and females for the calibration sample (Table 2) and validation dataset (Table 3). Considering this all together, they reveal a clear separation by sex using principal component analysis (Fig. 13), thus demonstrating their joint discriminant power. Furthermore, the measurements present similar correlation patterns for both sexes (Fig. 14). Some variables, such as SPU or IIMT, have only weak correlations with all other variables. Although a few pairs of variables exhibit correlation coefficients around 0.75, there is no clear pattern of strong multicollinearity: the classical problem of intercorrelation is avoided here, thanks to the choice of measurements, thus allowing for a greater efficiency in LDA when combining multiple variables.

Statistical processing of innominate data

DSP is based on Fisher's LDA, a classical and robust method of statistical classification. Historically widely used in the field of biological anthropology, it provides fast and accurate results, without requiring any intervention from the user on the selection of predictors.

There are two possible approaches for LDA, namely a geometrical approach and a Bayesian one. When the data do not strongly violate the hypotheses of multinormality and equality of covariance matrices, and when the prior probabilities are equal for all classes—which is the case with DSP—these two approaches are strictly equivalent (Saporta, 2011). The posterior probabilities for a given individual x (which can be assimilated to the real vector composed by its nonmissing anatomical measurements) can then be derived using the following formulae:

$$p_{\text{Male}}(x) = \frac{1}{1 + \exp\left(0.5 \times \left(d^2_{\text{Intra}}(x, G_{\text{Male}}) - d^2_{\text{Intra}}(x, G_{\text{Female}})\right)\right)}$$

$$p_{\text{Female}}(x) = 1 - p_{\text{Male}}(x)$$

Table 2 Descriptive statistics (by sex) for innominate variables in the metapopulation calibration (reference) sample.

	Males					**Females**					
Mes.	***n***	**Mean**	**SD**	**Min**	**Max**	***n***	**Mean**	**SD**	**Min**	**Max**	***p*-values**
PUM	992	69.7	5.2	56	85.8	988	72.3	5.2	57	86.8	<2e-16
SPU	1009	29.4	2.9	20.7	38.2	1013	24.4	2.6	17	32	<2e-16
DCOX	1010	213.4	13.3	173	253	1010	197	11.3	170	226	<2e-16
IIMT	1014	39.8	5.6	23	57	1022	45.4	5.6	31	63	<2e-16
ISMM	1005	112.3	6.8	93	131.2	1011	100.9	5.5	86.9	117.8	<2e-16
SCOX	984	155.7	11.1	123	187	994	151.9	10.6	126	183	<2e-14
SS	1014	73.6	5.7	56.4	91	1021	67.3	4.8	52.2	80.8	<2e-16
SA	1006	74.3	6.3	55.1	93.2	1021	74.7	6.7	53.3	94.7	0.16
SIS	1014	39.3	4.1	28	52	1008	35.4	3.5	26.6	46.3	<2e-16
VEAC	1013	56.3	3.4	47.3	66.3	1014	50.8	3	42	59.7	<2e-16

n = sample size; SD = standard deviation. The last column gives the *P*-value of a *t*-test for the comparison of means between sexes.

Table 3 Descriptive statistics (by sex) for innominate variables in the validation samples.

		Males					**Females**					
	Mes.	***n***	**Mean**	**SD**	**Min**	**Max**	***n***	**Mean**	**SD**	**Min**	**Max**	***P*-values**
Geneva	PUM	105	72.8	4	62.3	82.5	92	75.1	4.7	65.1	87.4	3e-4
	SPU	168	29.6	2.2	20.7	35	137	24.5	2.6	19	34.3	<2.2e-16
	DCOX	181	222.8	10.1	194	249	173	202.9	9.7	181	225	<2.2e-16
	IIMT	246	41.9	5.1	27	58	244	47.6	5.3	36	64	<2.2e-16
	ISMM	200	115.8	5.7	96.7	131	176	101.8	5.1	86.4	115	<2.2e-16
	SCOX	145	162.5	8	143	188	140	157.9	8.7	140	179	4.4e-6
	SS	247	76.2	4.8	60.5	88	246	69.8	4.8	57.8	82.8	<2.2e-16
	SA	246	76.9	7.2	0	90.5	246	78.2	6	63.4	96.9	0.03
	SIS	241	40.6	3.2	30.5	49.5	236	37	3.7	26.1	49.8	<2.2e-16
	VEAC	242	58.3	3.1	49.5	67.3	237	51	2.9	43.7	60.7	<2.2e-16
Maxwell	PUM	60	72	5.3	61.1	81	53	72.5	4.8	59.4	81.2	0.65
	SPU	60	30.1	2.9	24.9	36.3	58	25.1	1.9	21.7	30.1	<2.2e-16
	DCOX	61	220.8	13.1	195	242	59	198.7	10.6	169	217	<2.2e-16
	IIMT	59	42.1	5.3	29	53	59	49	5.8	39	66	4.6e-10
	ISMM	60	114.8	6.9	102.5	129	59	100	5.3	81.5	109.4	<2.2e-16
	SCOX	60	160.3	8.9	140	175	58	153.9	8.8	125	169	1e-4
	SS	59	77.2	6.2	63.9	88.5	59	70.4	4.7	59	89	1e-9
	SA	60	79.5	6.7	65.8	92.8	59	78.6	7.9	59.8	96.3	0.53
	SIS	59	41.3	3.4	33.5	47.3	57	37	2.8	29.1	43.7	1e-11
	VEAC	60	57.8	3.2	51.3	65.9	59	50.7	2.7	45	57.4	<2.2e-16

n = sample size; SD = standard deviation. The last column gives the *P*-value of a *t*-test for the comparison of means between sexes.

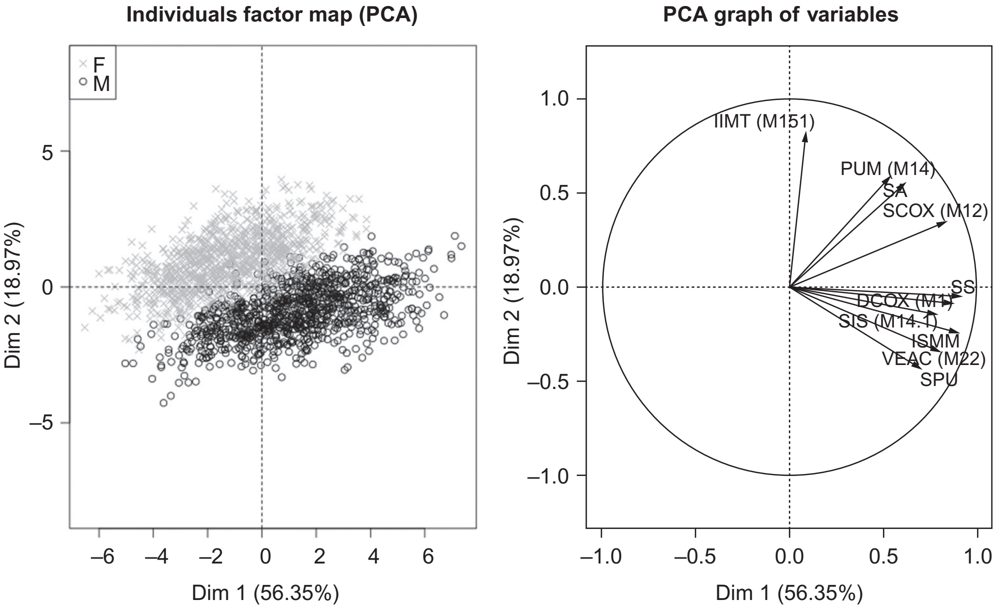


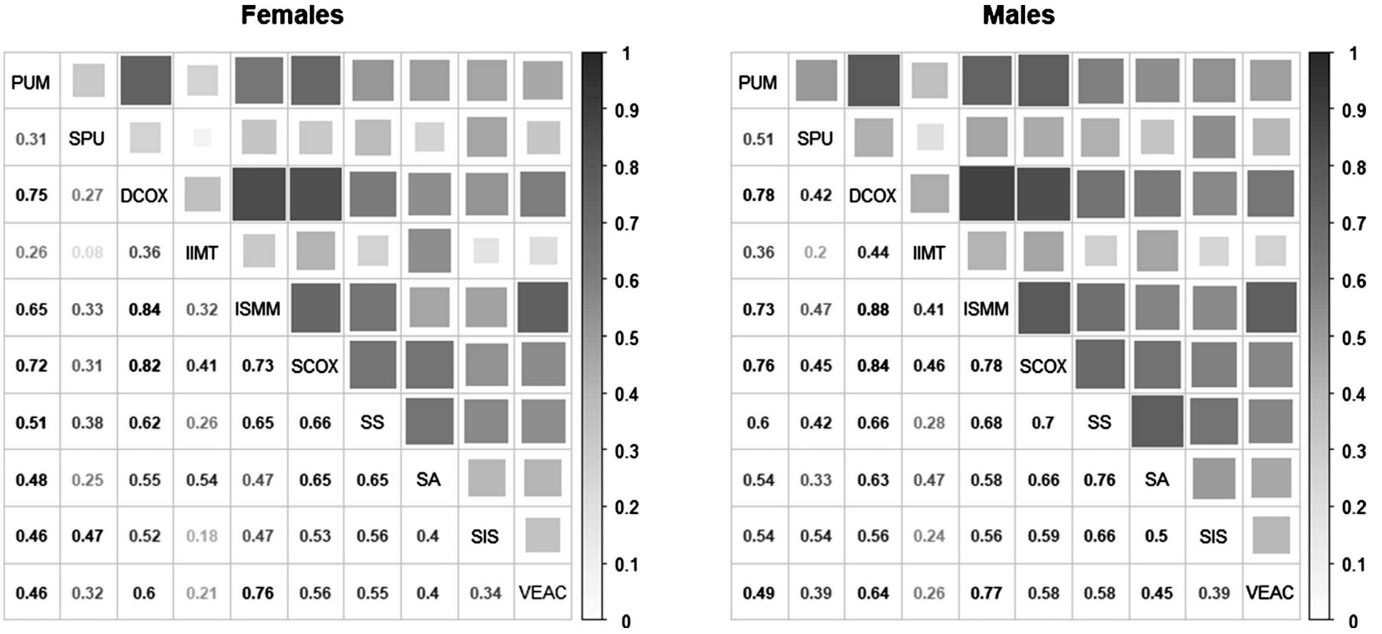
Fig. 13 Principal component analysis for the 2040 innominates described by the 10 measurements considered by DSP. 1866 innominates had all measurements available, whereas missing values were imputed using a regularized iterative algorithm for the 174 other ones (Josse & Husson, 2013).

where G_{Female} is the centroid of the female group for the corresponding variables, G_{Male} is the centroid of the male group, and d^2_{Intra} is the Mahalanobis distance associated with the intraclass covariance matrix (Bardos, 2001).

In the usual framework of LDA, an individual would be assigned to the group for which they would obtain the maximal posterior probability, i.e., an individual would be identified as male if $p_{\text{Male}} > 0.5$, and female otherwise. However, for more reliable sex estimations, DSP conforms to the conservative decision rule adopted in osteological studies: a posterior probability of 0.95 is considered a safe classification threshold (Kranioti & Apostol, 2015). Any individual who does not reach this value will remain indeterminate.

DSP2 software

At least 4 out of 10 variables are required to estimate a sex using DSP, which is supposed to be the minimum number of variables required to capture a reasonable amount of information on pelvic shape. Consequently, there are a total of $\sum_{i=4}^{10} \binom{10}{i} = 848$ possible combinations of variables that can be used as inputs in DSP. Publishing an exhaustive list of discriminant functions for all combinations would not be practical; the provision of a software implementing the method is preferred.



Females

PUM									
0.31	SPU								
0.75	0.27	DCOX							
0.26	0.08	0.36	IIMT						
0.65	0.33	0.84	0.32	ISMM					
0.72	0.31	0.82	0.41	0.73	SCOX				
0.51	0.38	0.62	0.26	0.65	0.66	SS			
0.48	0.25	0.55	0.54	0.47	0.65	0.65	SA		
0.46	0.47	0.52	0.18	0.47	0.53	0.56	0.4	SIS	
0.46	0.32	0.6	0.21	0.76	0.56	0.55	0.4	0.34	VEAC

Males

PUM									
0.51	SPU								
0.78	0.42	DCOX							
0.36	0.2	0.44	IIMT						
0.73	0.47	0.88	0.41	ISMM					
0.76	0.45	0.84	0.46	0.78	SCOX				
0.6	0.42	0.66	0.28	0.68	0.7	SS			
0.54	0.33	0.63	0.47	0.58	0.66	0.76	SA		
0.54	0.54	0.56	0.24	0.56	0.59	0.66	0.5	SIS	
0.49	0.39	0.64	0.26	0.77	0.58	0.58	0.45	0.39	VEAC

Fig. 14 Visualization of the correlation matrix calculated on the 1866 complete innominates from the calibration sample used by DSP. The correlation coefficients are classical Pearson linear coefficients.

In 2005, DSP was originally published with an Excel spreadsheet created by Francis Houët. Although still available (accessible at http://projets.pacea.u-bordeaux.fr/logiciel/DSP/dsp.xls), it should no longer be used since some compatibility issues arise with more recent versions of Excel. That is why a free and cross-platform software, DSP2, has been made available in 2017. Coded in C++ and using the Qt library for the graphical user interface—programmed by Bruno Dutailly, CNRS engineer at the PACEA laboratory, University of Bordeaux, France—it also uses a portable version of the R statistical software. Data can be entered manually in DSP2 or copy-pasted from a previously created spreadsheet. DSP2 also checks the general integrity of the data submitted by the user: to avoid any erroneous measurement or data entry mistakes, the data are automatically compared to the range of calibration dataset, and any suspicious value is indicated by an orange color. The range of acceptable values for each variable can also be directly consulted in the software.

DSP2 can be freely downloaded from the following website: http://projets.pacea.u-bordeaux.fr/logiciel/DSP2/dsp2.html, and needs no registration. The graphical user interface of DSP2 is presented in Fig. 15. As an alternative, and for an easier use on mobile devices, an online R-shiny application implementing the algorithm of DSP has also been developed independently by João Coelho and David Navega (University of Coimbra). This application is available on their website dedicated to various anthropological methods, Osteomics (http://apps.osteomics.com/DSP/).

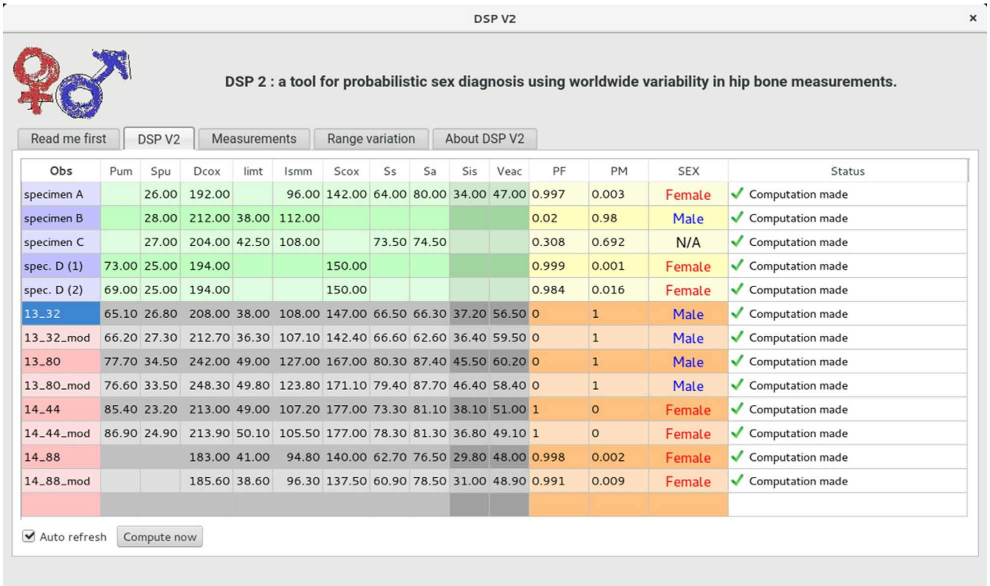
Study design

In the original publication of DSP, Murail et al. (2005) evaluated the accuracy obtained with 4 out of 848 possible combinations of variables:

1. All 10 variables.
2. A combination of eight variables that excluded SIS and VEAC, those two being rather considered as "rescue" variables, mainly used for incomplete bones.
3. The best combination of four variables according to Wilk's lambda value in LDA: DCOX, PUM, SPU, and IIMT.
4. And the worst combination of four variables according to Wilk's lambda value: SIS, VEAC, SS, and SA.

Here, two new combinations of variables are also tested to further evaluate the applicability of DSP:

5. A subset of six variables: DCOX, SCOX, SS, SA, SIS, and VEAC. This subset includes the worst combination of four variables and is not expected to deliver the best possible accuracy, but it fits some practical considerations, excluding PUM (often badly preserved on ancient material) and IIMT (more likely to be subject to measurement error).



Obs	Pum	Spu	Dcox	Iimt	Ismm	Scox	Ss	Sa	Sis	Veac	PF	PM	SEX	Status
specimen A		26.00	192.00		96.00	142.00	64.00	80.00	34.00	47.00	0.997	0.003	Female	✓ Computation made
specimen B		28.00	212.00	38.00	112.00						0.02	0.98	Male	✓ Computation made
specimen C		27.00	204.00	42.50	108.00		73.50	74.50			0.308	0.692	N/A	✓ Computation made
spec. D (1)	73.00	25.00	194.00			150.00					0.999	0.001	Female	✓ Computation made
spec. D (2)	69.00	25.00	194.00			150.00					0.984	0.016	Female	✓ Computation made
13_32	65.10	26.80	208.00	38.00	108.00	147.00	66.50	66.30	37.20	56.50	0	1	Male	✓ Computation made
13_32_mod	66.20	27.30	212.70	36.30	107.10	142.40	66.60	62.60	36.40	59.50	0	1	Male	✓ Computation made
13_80	77.70	34.50	242.00	49.00	127.00	167.00	80.30	87.40	45.50	60.20	0	1	Male	✓ Computation made
13_80_mod	76.60	33.50	248.30	49.80	123.80	171.10	79.40	87.70	46.40	58.40	0	1	Male	✓ Computation made
14_44	85.40	23.20	213.00	49.00	107.20	177.00	73.30	81.10	38.10	51.00	1	0	Female	✓ Computation made
14_44_mod	86.90	24.90	213.90	50.10	105.50	177.00	78.30	81.30	36.80	49.10	1	0	Female	✓ Computation made
14_88			183.00	41.00	94.80	140.00	62.70	76.50	29.80	48.00	0.998	0.002	Female	✓ Computation made
14_88_mod			185.60	38.60	96.30	137.50	60.90	78.50	31.00	48.90	0.991	0.009	Female	✓ Computation made

Fig. 15 Graphical user interface of DSP2 software. The first five lines are the native examples given when opening the software. The last eight lines correspond to two individuals randomly extracted from the Maxwell collection (13_32 and 13_80) and two randomly extracted from the Simon collection (14_44 and 14_88). For their modified version ("_mod"), we applied a random Gaussian noise to their measurements to evaluate the robustness of DSP2 to measurement errors.

6. A subset of five variables: PUM, ISMM, SS, SA, VEAC. This subset excludes the two measurements of os coxae completely (DCOX and SCOX).

For all these combinations, learning error on the calibration dataset and validation error on the target dataset are evaluated.

Results

Sex classification accuracy

Table 4 presents the learning error obtained by DSP2 on the calibration dataset, using various combinations of variables as predictors. In all the cases, the learning error remains below 1.4%, even with the "worst" possible combination of variables. Conversely, the percentage of individuals whose sex can be estimated (i.e., reaching a posterior probability >0.95, regardless of whether the sex estimate is correct or not) strongly depends on the number and, above all, the nature of the variables used for estimation. Using the best eight variables, >90% of individuals can be determined,

Table 4 Results with various combinations of variables (vars) for the calibration sample ("% sexing": percentage of innominate whose sex could be estimated, i.e., reaching the classification threshold; "% accuracy": learning accuracy, i.e., percentage of innominates correctly determined among those that reach the classification threshold).

	Whole sample		Females		Males	
	% sexing	% accuracy	% sexing	% accuracy	% sexing	% accuracy
All 10 vars	90.8	99.7	91	99.3	90.7	100
8 vars (without SIS and VEAC)	91	99.7	91.2	99.3	90.8	100
6 vars (DCOX, SCOX, SS, SA, SIS, VEAC)	58.2	99.2	59.9	99.5	56.6	98.9
5 vars (PUM, ISMM, SS, SA, VEAC)	76.3	99	75.8	99	76.7	99
Best 4 vars (DCOX, PUM, SPU, IIMT)	87.2	99.5	87.6	99.5	86.8	99.5
Worst 4 vars (SIS, VEAC, SA, SS)	41.5	98.7	42.5	98.8	40.4	98.5

but using the worst combination of four variables, almost 60% on individuals remain indeterminate.

It should be noted that the best combination of four variables outperforms the studied combination of six variables in both percentage of accuracy and percentage of estimation. Using a few strong predictors can then be more concluding than retaining a greater number of variables from the acetabular and sacro-iliac modules, whose discriminant power is generally lower.

Sex classification reliability

Table 5 presents the detailed results for each collection of the calibration sample, in search for geographical differences in the efficiency of DSP2. Even if a moderate variability in sexing rates can be noted among the 12 population samples, the difference in reliability rates is negligible regardless of the combination of variables: in the worst case (which happens for the worst combination of four variables), the coefficient of variation of reliability rates among the 12 population samples is only 1.5%.

The performance of DSP2 on the validation sample (i.e., the reliability of the method) is also evaluated with the Maxwell and Simon collections. Depending on the combination of variables retained, DSP2 reaches a sex diagnosis for 50%–95% of individuals. The error rate reaches a maximum of 5% in the least favorable scenario, but is globally around 1% for Maxwell collection (Table 5, #13), and between 2% and 4% for Simon collection (Table 5, #14).

Table 5 Results with various combinations of variables (vars) for the calibration sample and validation sample ("% sexing": percentage of innominates whose sex could be estimated, i.e., reaching the classification threshold; "% reliability": percentage of innominates correctly determined among those that reach the classification threshold. The accuracy for the reference sample is the learning error; the reliability for the validation sample is the prediction error. For details about the combinations of variables, see Table 4).

	10 vars		8 vars		6 vars		5 vars		Best 4 vars		Worst 4 vars	
Sample	**% sexing**	**% reliability**	**% sexing**	**% reliability**	**% sexing**	**% reliability**	**% sexing**	**% reliability**	**% sexing**	**% reliability**	**% sexing**	**% reliability**
1	89.9	100	90.1	100	52	100	73.8	100	83.2	99.2	45.6	100
2	89.6	100	89.7	100	65.1	97.8	83.1	99.4	86.5	98.3	44.8	99
3	86.5	100	86.5	100	66.7	100	83	100	80.7	100	32.8	100
4	95.4	100	95.4	100	66.1	100	83.6	100	92.7	100	39.9	100
5	88.4	99.2	88.8	99.2	47.7	98.6	67.8	98	84.9	100	33.4	98
6	85.4	100	85.4	100	47.1	100	62.6	97	80.7	100	43.4	95.7
7	93.6	100	94	100	60.2	100	85.4	98.9	89.7	100	43.3	100
8	93.6	100	93.7	100	61.9	100	78.1	100	90.5	100	42.4	100
9	90.2	98.9	90.2	98.9	54.7	100	76	97.5	87	100	40.2	97.8
10	88.2	100	89	100	66.3	98.4	80.1	99.4	84	99.4	49.2	97.9
11	91.7	98.4	91.4	98.4	60.5	98.4	70.5	99.3	89.6	98.4	45	96.9
12	94.6	100	94.6	100	55.5	100	78.2	98.6	91.1	100	38.1	100
13	93.5	99	93.6	99	62.3	100	82.7	98.9	87.3	99	50.9	94.9
14	94.8	96.1	94.8	96.1	71.6	97.7	85	95	90.9	96.9	55.3	98.1

Additionally, for both calibration and validation datasets, the difference in reliability rates for males and females is also negligible (always inferior to 0.8% for all six combinations of variables tested; results not shown on Table 5): the decision rule of DSP2 includes no bias and predicts both classes with the same efficiency.

Robustness to measurement uncertainty

Since some of the measurements used in DSP2 may be deemed complex to record without prior training, measurement uncertainty on sex estimates could have an impact in some practice. In order to evaluate this uncertainty, four individuals (two of them in Maxwell collection, two others in Simon collection) were randomly chosen within the validation sample. All four individuals were correctly classified with a posterior probability >0.998. To simulate the effect of measurement errors, we applied to each value a centered Gaussian noise, whose standard deviation was set to 5% of the corresponding measurement. This deviation of 5% was chosen to represent a reasonable measurement error in anthropometry for a moderately experienced user equipped with a caliper. Those "modified" versions of the four individuals were then submitted to DSP2, and all of them were still correctly classified with a posterior probability >0.991 (Fig. 13). This suggests that DSP2 remains reliable in the presence of a moderate measurement error.

Discussion and conclusions

Sex estimation of unknown human remains is a subject increasingly investigated in the biological anthropology community (e.g., Dirkmaat, 2014; Langley & Tersigni-Tarrant, 2017; Nikita, 2016). Whether the practitioner aims at a more objective analysis of a past population, or at an accurate diagnosis of a forensic case using a method that respects the *Daubert* criteria (e.g., Klales & Burns, 2017; Kotěrová et al., 2017), there has been a shift from the interest on the success rate of a technique to a zero-tolerance policy toward the error rate. Using the pelvic morphology for sex estimation ensures a low error rate. Furthermore, introducing the possibility to accept an individual as indeterminate sex allows for lowering this error rate.

The control of error rate is of utmost importance in forensic cases (Christensen, Crowder, Ousley, & Houck, 2014), and DSP offers a robust and cost-efficient technique to estimate sex when the innominate is preserved. We have met in DSP the criterion for a reliable classification consisting in a sectioning point of 0.95 for posterior probability (Kranioti & Apostol, 2015). This allows to reduce drastically the classification error rate compared to the conventional threshold of 0.50 used in discriminant function analyses, which should not be used in biological anthropology and forensic sciences (Kazzazi & Kranioti, 2018). When the sectioning point for a correct classification of 0.50 is employed, there is a large number of individuals in the overlapping area with similar posterior probability values—that is, with nearly equal chances of being a male or a female.

After Berg (2017), a good discriminant function produces a cross-validated accuracy rate between 85% and 95%. But this does not mean that a risk of error up to 15% can be acceptable in all contexts, particularly in forensic anthropology cases. It cannot be forgotten that the overlap of both sexes is much higher than these values and is dependent on variability.

DSP shows accuracy close to perfection, using more than 2000 innominates from around the world. The versatility of the method allows for the use of different sets of variables among the 10 included in the model. The success of the method is ensured by the use of these specific measurements that closely reflect the morphology that is sexually dimorphic in the whole human species.

As stated by Berg and Kenyhercz (2017), examples of sex estimation using quantitative measures are proliferative in the anthropological literature, and because of this abundance, only a few examples are discussed in this text, focused on the probabilistic approach. From this point of view, it can be reminded (Komar & Buikstra, 2008) that the more universally tested the models are on global databases, such as FORDISC (Ousley & Jantz, 2012) (see Chapter 12) and DSP (Brůžek et al., 2017; Murail et al., 2005), the more robust and reliable the sex estimates they offer can be considered. DSP has recently gained attention from different parts of the worldwide community (e.g., Krishan et al., 2016; Machado et al., 2018; Moore, DiGangi, Ruíz, Davila, & Medina, 2016; Spradley, 2016). It has been used in a number of studies in bioarchaeology and forensic anthropology contexts (e.g., Baker et al., 2017; Candelas González, Rascón Pérez, Chamero, Cambra-Moo, & González Martín, 2017; Hansen et al., 2017; Oelze et al., 2012; Quintelier, 2009; Ríos, García-Rubio, Martínez, Alonso, & Puente, 2012; Talamo et al., 2018; Villotte, Santos, & Courtaud, 2015). Recently, Jerković, Bašić, Kružić, and Anđelinović (2018) tested the applicability of DSP in a bioarchaeological context through validation on known sex data obtained by DNA analyses on a medieval sample. They recommend the implementation of DSP for creating calibration data and development of metric and nonmetric population-specific sex estimation standards in past populations.

In forensic anthropology, *Daubert* rules require an extensive validation of the method (Christensen & Crowder, 2009). DSP has recently been tested in different population samples of dry bones and models derived from CT imaging. The authors have confirmed a high reliability: with the exception of two individuals (Chapman et al., 2014), no classification error was found among 206 individuals tested (Mestekova, Brůžek, Veleminska, & Chaumoitre, 2015; Quatrehomme, Radoman, Nogueira, du Jardin, & Alunni, 2017).

Although DSP is applicable to several populations, few potential exceptions remain to be further investigated: a validation on a Mexican population proved to be error-free, but with a relatively high number of indeterminate individuals for the male subsample (Sánchez-Mejorada, Gómez-Valdés, Herrera, Veleminsky, & Brůžek, 2011) and a number of misclassifications in a Brazilian population (Machado et al., 2018). From the 103 innominate analyzed, there was 9.4% of error in classification for Brazilian male individuals, and 14% for females. It is interesting to note that in this study the highest accuracy has been achieved using the four worst variables, instead of the combination of best variables. Any attempt to explain these results is pure speculation at this stage. Only a new study on a Brazilian sample, which can be expected soon, may explain these findings.

More than a decade after the first paper on DSP (Murail et al., 2005), it is time to improve and increase its use, which is the reason why we are responding to some criticisms herein (e.g., Baumgarten & Ousley, 2015; Baumgarten, Ousley, Decker, & Shirley, 2015). The authors expressed some concerns about the replication of landmarks and considered that DSP did not provide a typical logistic regression (i.e., with a classification rule). These two studies achieved similar accuracy compared to the results presented in Murail et al. (2005). Controversial details have been now clarified by Brůžek et al. (2017) and the present chapter.

To conclude, DSP2 is a population nonspecific, well-defined, user-friendly, robust, and reliable technique for sexing the innominate in forensic anthropology, as well as in bioarchaeology that conforms to the *Daubert* standards.

References

Adams, B. J., & Byrd, J. E. (2002). Interobserver variation of selected postcranial skeletal measurements. *Journal of Forensic Science*, *47*(6), 1193–1202.

Ahn, A. C., Tewari, M., Poon, C. S., & Phillips, R. S. (2006). The limits of reductionism in medicine: Could systems biology offer an alternative? *PLoS Medicine*, *3*(6), e208.

Baker, O., Chamel, B., Coqueugniot, É., Khawam, R., Stordeur, D., Perrin, P., et al. (2017). Prehistory of human tuberculosis: Earliest evidence from the onset of animal husbandry in the Near East. *Paléorient*, *43*, 35–51.

Bardos, M. (2001). *Analyse discriminante*. Dunod.

Barthélémy, I. (1999). *Morphologie et évolution: le peuplement du Sud-Ouest de la France entre le VIème et le XIIème siècles (le cas des nécropoles de l'Islejourdain-La Gravette, Gers)*. PhD thesis Toulouse, France: Université de Toulouse 3.

Baumgarten, S. E., & Ousley, S. D. (2015). Estimating sex from the innominate utilizing new measurements. In *Vol. 21. Proceedings of the 67th Annual Scientific Meeting AAFS, February 16–21, Orlando, FL* (p. 142).

Baumgarten, S. E., Ousley, S. D., Decker, S. J., & Shirley, N. R. (2015). A transparent method for sex estimation using refined DSP measurements of the innominate. *American Journal of Physical Anthropology*, *156*, 80.

Berg, G. E. (2017). Sex estimation of unknown human skeletal remains. In N. R. Langley & M. A. Tersigni-Tarrant (Eds.), *Forensic anthropology: A comprehensive introduction* (pp. 143–162). Boca Raton: CRC Press.

Berg, G. E., & Kenyhercz, M. W. (2017). Introducing human mandible identification [(hu) MANid]: A free, web-based GUI to classify human mandibles. *Journal of Forensic Sciences*, *62*(6), 1592–1598.

Bräuer, G. (1988). Osteometrie. In R. Knussmann (Ed.), *Anthropologie, handbuch der vergleichenden Biologie des Menschen, Band 1* (pp. 160–232). Stuttgart: G. Fischer Verlag.

Brůžek, J. (1984). *Dimorphisme sexuel de l'os coxal de l'Homme du point de vue ontogénique et phylogénique*. PhD Thesis Prague, Czech Republic: Charles University.

Brůžek, J. (1991). *Fiabilité des procédés de détermination du sexe à partir de l'os coxal, Implications à l'étude du dimorphisme sexuel de l'Homme fossile*. PhD Thesis Paris: Muséum National d'Histoire Naturelle, Institut de Paléontologie Humaine. 431 pp.

Brůžek, J. (1992). Fiabilité des fonctions discriminantes dans la détermination sexuelle de l'os coxal. Critiques et propositions. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, *4*(1), 67–104.

Brůžek, J., Murail, P., Houët, F., & Cleuvenot, E. (1994). Inter- and intra-observer error in pelvic measurements and its application for the methods of sex determination. *Anthropologie (Brno)*, *32*, 215–223.

Brůžek, J., Santos, F., Dutailly, B., Murail, P., & Cunha, E. (2017). Validation and reliability of the sex estimation of the human os coxae using freely available DSP2 software for bioarchaeology and forensic anthropology. *American Journal of Physical Anthropology*, *164*(2), 440–449.

Buikstra, J. E., & Ubelaker, D. H. (1994). *Standards for data collection from human skeletal remains*. In *Proceedings of a seminar at the Field Museum of Natural History, organized by Jonathan Haas* (3rd ed.). Fayetteville, Arkansas: Arkansas Archeological Survey.

Byers, S. N. (2002). *Introduction to forensic anthropology: A textbook*. Boston: Allyn and Bacon.

Cameron, N. (1986). The methods of auxological anthropometry. In F. Falkner & J. M. Tanner (Eds.), *Vol. 3. Human growth—A comprehensive treatise: Methodology, ecological, genetic and nutritional effects on growth* (pp. 3–46). New York: Plenum Press.

Candelas González, N., Rascón Pérez, J., Chamero, B., Cambra-Moo, O., & González Martín, A. (2017). Geometric morphometrics reveals restrictions on the shape of the female os coxae. *Journal of Anatomy*, *230*(1), 66–74.

Chapman, T., Lefevre, P., Semal, P., Moiseev, F., Sholukha, V., Louryan, S., et al. (2014). Sex determination using the probabilistic sex diagnosis (DSP: Diagnose Sexuelle Probabiliste) tool in a virtual environment. *Forensic Science International*, *234*. 189.e1–189.e8.

Christensen, A. M., & Crowder, C. M. (2009). Evidentiary standards for forensic anthropology. *Journal of Forensic Sciences*, *54*(6), 1211–1216.

Christensen, A. M., Crowder, C. M., Ousley, S. D., & Houck, M. M. (2014). Error and its meaning in forensic science. *Journal of Forensic Sciences*, *59*(1), 123–126.

Davivongs, V. (1963). The pelvic girdle of the Australian aborigine; sex differences and sex determination. *American Journal of Physical Anthropology*, *21*(4), 443–455.

Derry, D. E. (1923). On the sexual and racial characters of the human ilium. *Journal of Anatomy*, *58*(1), 71.

Dirkmaat, D. (2014). *A companion to forensic anthropology*. John Wiley & Sons.

Drew, R. (2013). A review of the Ischium-Pubis index: Accuracy, reliability, and common errors. *Human Biology*, *85*(4), 579–596.

Dwight, T. (1878). *The identification of the human skeleton*. Boston: David Clapp & Son.

Esteve-Altava, B. (2017a). Challenges in identifying and interpreting organizational modules in morphology. *Journal of Morphology*, *278*(7), 960–974.

Esteve-Altava, B. (2017b). In search of morphological modules: A systematic review. *Biological Reviews*, *92*(3), 1332–1347.

Frayer, D. W., & Wolpoff, M. H. (1985). Sexual dimorphism. *Annual Review of Anthropology*, *14*(1), 429–473.

Gaillard, J. (1960). Détermination sexuelle d'un os coxal fragmentaire. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, *1*(2), 255–267.

Gaillard, J. (1961). Valeur de l'indice ischio-pubien pour la détermination sexuelle de l'os coxal. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, *2*(1), 92–108.

Genovés, S. (1959). L'estimation des différences sexuelles dans l'os coxal: Différences métriques et différences morphologiques. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, *10*(1), 3–95.

Grabowski, M. W., Polk, J. D., & Roseman, C. C. (2011). Divergent patterns of integration and reduced constraint in the human hip and the origins of bipedalism. *Evolution: International Journal of Organic Evolution*, *65*(5), 1336–1356.

Hanna, R. E., & Washburn, S. L. (1953). The determination of the sex of skeletons as illustrated by a study of the Eskimo pelvis. *Human Biology*, *25*(1), 21.

Hansen, J., Petersen, H. C., Frei, K. M., Courtaud, P., Tillier, A. M., Fischer, A., et al. (2017). The Maglemosian skeleton from Koelbjerg, Denmark revisited: Identifying sex and provenance. *Danish Journal of Archaeology*, *6*(1), 50–66.

Henderson, C. Y., Mariotti, V., Pany-Kucera, D., Villotte, S., & Wilczak, C. (2013). Recording specific entheseal changes of fibrocartilaginous entheses: Initial tests using the Coimbra method. *International Journal of Osteoarchaeology*, *23*(2), 152–162.

Hoyme, L. E. (1957). The earliest use of indices for sexing pelves. *American Journal of Physical Anthropology*, *15*(4), 537–546.

İşcan, M. Y., & Steyn, M. (2013). *The human skeleton in forensic medicine*. Charles C Thomas Publisher.

Jerković, I., Bašić, Ž., Kružić, I., & Anđelinović, Š. (2018). Creating reference data on sex for ancient populations using the Probabilistic Sex Diagnosis method: A validation test using the results of aDNA analysis. *Journal of Archaeological Science*, *94*, 44–50.

Josse, J., & Husson, F. (2013). Handling missing values in exploratory multivariate data analysis methods. *Journal de la Société française de statistique*, *153*(2), 79–99.

Jovanović, S., & Živanović, S. (1965). The establishment of the sex by the great schiatic notch. *Cells Tissues Organs*, *61*(1), 101–107.

Kazzazi, S. M., & Kranioti, E. F. (2018). Sex estimation using cervical dental measurements in an archaeological population from Iran. *Archaeological and Anthropological Sciences*, *10*(2), 439–448.

Kim, D. H., Lee, S. S., Han, S. H., & Lee, U. Y. (2018). A new landmark for measuring the Ischium-Pubis index for sex determination by using three-dimensional models of South Korean population. *Australian Journal of Forensic Sciences*, *50*(5), 472–481.

Kimura, K. (1982). Sex differences of the hip bone among several populations. *Okajimas Folia Anatomica Japonica*, *58*(4-6), 265–275.

Klales, A. R., & Burns, T. L. (2017). Adapting and applying the Phenice (1969) adult morphological sex estimation technique to subadults. *Journal of Forensic Sciences*, *62*(3), 747–752.

Klingenberg, C. P. (2008). Morphological integration and developmental modularity. *Annual Review of Ecology, Evolution, and Systematics*, *39*(1), 115–132.

Komar, D. A., & Buikstra, J. E. (2008). *Forensic anthropology: Contemporary theory and practice*. New York: Oxford university press.

Komar, D. A., & Grivas, C. (2008). Manufactured populations: What do contemporary reference skeletal collections represent? A comparative study using the Maxwell Museum documented collection. *American Journal of Physical Anthropology*, *137*(2), 224–233.

Kotěrová, A., Velemínská, J., Dupej, J., Brzobohatá, H., Pilný, A., & Brůžek, J. (2017). Disregarding population specificity: Its influence on the sex assessment methods from the tibia. *International Journal of Legal Medicine*, *131*(1), 251–261.

Kranioti, E. F., & Apostol, M. A. (2015). Sexual dimorphism of the tibia in contemporary Greeks, Italians, and Spanish: Forensic implications. *International Journal of Legal Medicine*, *129*(2), 357–363.

Krishan, K., Chatterjee, P. M., Kanchan, T., Kaur, S., Baryah, N., & Singh, R. K. (2016). A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. *Forensic Science International*, *165*. 165.e1–165.e8.

Langley, N. R., Jantz, L. M., McNulty, S., Maijanen, H., Ousley, S. D., & Jantz, R. L. (2018). Error quantification of osteometric data in forensic anthropology. *Forensic Science International*, *287*, 183–189.

Langley, N. R., & Tersigni-Tarrant, M. A. (Eds.), (2017). *Forensic anthropology: A comprehensive introduction*: CRC Press.

Lazorthes, G., & Lhès, A. (1939). La grande échancrure sciatique: Étude de la morphologie et de ses caractères sexuels. *Archives d'anatomie, d'histologie et d'embryologie normales et expérimentales, Strasbourg*, *27*, 143–170.

Letterman, G. S. (1941). The greater sciatic notch in American whites and Negroes. *American Journal of Physical Anthropology*, *28*(1), 99–116.

Lewton, K. L. (2012). Evolvability of the primate pelvic girdle. *Evolutionary Biology*, *39*(1), 126–139.

Machado, M. P. S., Costa, S. T., Freire, A. R., Navega, D., Cunha, E., Júnior, E. D., et al. (2018). Application and validation of Diagnose Sexuelle Probabiliste V2 tool in a miscegenated population. *Forensic Science International*, *290*, 351.e1–351.e5.

Mestekova, S., Brůžek, J., Veleminska, J., & Chaumoitre, K. (2015). A test of the DSP sexing method on CT images from a modern French sample. *Journal of Forensic Sciences*, *60*(5), 1295–1299.

Moeschler, P. (1964). *Structures morphologiques et dimorphisme sexuel: Essai de différenciation métrique: Application à l'os coxal*. PhD thesis Switzerland: University of Geneva.

Moore, M. K., DiGangi, E. A., Ruíz, F. P. N., Davila, O. J. H., & Medina, C. S. (2016). Metric sex estimation from the postcranial skeleton for the Colombian population. *Forensic Science International*, *262*. 286.e1–286.e8.

Murail, P., Brůžek, J., Houët, F., & Cunha, E. (2005). DSP: A tool for probabilistic sex diagnosis using worldwide variability in hip-bone measurements. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, *17*(3–4), 167–176.

Nikita, E. (2016). *Osteoarchaeology: A guide to the macroscopic study of human skeletal remains*. Academic Press.

Novotný, V. (1981). *Pohlavnı rozdıly a identifikace pohlavá pınevnı kosti [Sex differences and identification of sex in pelvic bone]*. Doctoral dissertation, Ph. D. thesis Brno: Purkyne University.

Novotný, V. (1986). Sex determination of the pelvic bone: A systems approach. *Anthropologie*, *1962*, 197–206.

Oelze, V. M., Koch, J. K., Kupke, K., Nehlich, O., Zäuner, S., Wahl, J., et al. (2012). Multi-isotopic analysis reveals individual mobility and diet at the early iron age monumental tumulus of Magdalenenberg, Germany. *American Journal of Physical Anthropology*, *148*(3), 406–421.

Ousley, S. D., & Jantz, R. L. (2012). Fordisc 3 and statistical methods for estimating sex and ancestry. In D. C. Dirkmaat (Ed.), *A companion to forensic anthropology* (pp. 311–329). Chichester, UK: John Wiley & Sons, Ltd. [chap. 15].

Perréard-Lopreno, G. (2007). *Adaptation structurelle des os du membre supérieur et de la clavicule à l'activité: Analyse de l'asymétrie des propriétés géométriques de sections transverses et de mesures linéaires dans une population identifiée (collection Simon)*. PhD thesis Switzerland: Université de Genève.

Quatrehomme, G., Radoman, I., Nogueira, L., du Jardin, P., & Alunni, V. (2017). Sex determination using the DSP (probabilistic sex diagnosis) method on the coxal bone: Efficiency of method according to number of available variables. *Forensic Science International*, *272*, 190–193.

Quintelier, K. (2009). Calcified uterine leiomyomata from a post-medieval nunnery in Brussels, Belgium. *International Journal of Osteoarchaeology*, *19*(3), 436–442.

Ríos, L., García-Rubio, A., Martínez, B., Alonso, A., & Puente, J. (2012). Identification process in mass graves from the Spanish Civil War II. *Forensic Science International*, *219*(1–3), e4–e9.

Sánchez-Mejorada, G., Gómez-Valdés, J., Herrera, P., Veleminsky, P., & Brůžek, J. (2011). Valoración del método de diagnóstico sexual probabilístico (DSP) en una colección osteológica mexicana. *Estudios de Antropología Biológica*, *15*(1).

Saporta, G. (2011). *Probabilités, analyse des données et Statistique* (3rd ed.). Technip.

Schulter-Ellis, F. P., Schmidt, D. J., Hayek, L. C., & Craig, J. (1983). Determination of sex with a discriminant analysis of new pelvic bone measurements: Part I. *Journal of Forensic Sciences*, *28*(1), 169–180.

Schultz, A. H. (1930). The skeleton of the trunk and limbs of higher primates. *Human Biology*, *2*(3), 303–438.

Segebarth-Orban, R. (1980). An evaluation of the sexual dimorphism of the human innominate bone. *Journal of Human Evolution*, *9*(8), 601–607.

Seidler, H. (1980). Sex-diagnosis of isolated os coxae by discriminant functions. *Journal of Human Evolution*, *9*(8), 597–600.

Singh, S., & Potturi, B. R. (1978). Greater sciatic notch in sex determination. *Journal of Anatomy*, *125*(3), 619.

Spradley, M. K. (2016). Metric methods for the biological profile in forensic anthropology: Sex, ancestry, and stature. *Academic Forensic Pathology*, *6*(3), 391–399.

Straus, W. L., Jr. (1927). The human ilium: Sex and stock. *American Journal of Physical Anthropology*, *11*(1), 1–28.

Talamo, S., Samsel, M., Jaouen, K., Delvigne, V., Lafarge, A., Raynal, J. P., et al. (2018). A reassessment of the presumed Badegoulian skull from Rond-du-Barry cave (Polignac, France), using direct AMS radiocarbon dating. *American Journal of Physical Anthropology*.

Thieme, F. P., & Schull, W. J. (1957). Sex determination from the skeleton. *Human Biology*, *29*(3), 242.

Tin, N. (1938). *Recherches sur le bassin de la femme annamite*. PhD thesis Indochina: National Institute of Anatomy of the Obstetric Clinic of High School and Medicine and Pharmacy.

Utermohle, C. J., & Zegura, S. L. (1982). Intra- and interobserver error in craniometry: A cautionary tale. *American Journal of Physical Anthropology*, *57*, 303–310.

Vacca, E., & Di Vella, G. (2012). Metric characterization of the human coxal bone on a recent Italian sample and multivariate discriminant analysis to determine sex. *Forensic Science International*, *222*(1–3). 401.e1–401.e9.

Verneau, R. (1875). *Le bassin dans les sexes et dans les races*. JB Bailliere.

Villotte, S., Santos, F., & Courtaud, P. (2015). In situ study of the Gravettian individual from Cussac cave, locus 2 (Dordogne, France). *American Journal of Physical Anthropology*, *158*, 759–768.

von Bertalanffy, L. (1968). *General system theory* (p. 40). New York: George Braziller.

Waldron, T. (1987). The relative survival of the human skeleton: Implication for palaeopathology. In A. Boddindton, A. N. Garland, & R. C. Janaway (Eds.), *Death, decay and reconstruction* (pp. 55–64). Manchester: Manchester University Press.

Washburn, S. L. (1948). Sex differences in the pubic bone. *American Journal of Physical Anthropology*, *6*(2), 199–208.

Washburn, S. L. (1949). Sex differences in the pubic bone of Bantu and Bushman. *American Journal of Physical Anthropology*, 7(3), 425–432.

CHAPTER 16

MorphoPASSE: Morphological pelvis and skull sex estimation program

Alexandra R. Klales
Forensic Anthropology Program, Department of Sociology and Anthropology, Washburn University, Topeka, KS, United States

Background and rationale

The primary goal asked of forensic anthropologists is the estimation of a person's biological profile to help law enforcement determine the identity of unknown human remains encountered in forensic contexts. Sex estimation is an important component of the biological profile, and many methods have been developed, tested, and utilized to estimate sex using different regions of the human skeleton. Despite the availability of quantitative methods for sex estimation, many forensic anthropologists continue to rely on qualitative traits for sex estimation. In a survey of biological anthropologists, most practitioners (63.6%) prefer using both qualitative and quantitative methods; however, when both are not used, qualitative methods (23.9%) were preferred nearly twice as often as metric methods (12.5%) (Klales, 2013).

Benefits of morphological methods include ease of use, relatively quick application, no need for specialized equipment, and applicability to fragmentary remains. Unfortunately, many of the qualitative methods used for sex estimation are based on subjective interpretations of skull and pelvic traits (Rogers & Saunders, 1994; Williams & Rogers, 2006). Because of this, attempts to standardize the use of morphological traits has resulted in the creation of methods that rely on standardized ordinal scoring and statistical methods of sex classification (e.g., Klales, Ousley, & Vollner, 2012; Walker, 2008), thereby making the methods compliant with the *Daubert* recommendations (Daubert vs. Merrell Dow Pharmaceuticals, 1993). According to the Klales (2013) survey, the most preferred traits for morphological cranial sex estimation are the traits depicted in Buikstra and Ubelaker (1994) and subsequently utilized by Walker (2008): nuchal crest, mastoid process, supraorbital margin, glabella, and mental eminence. Likewise, survey participants indicated that the three traits originally described by Phenice (1969) and then modified by Klales et al. (2012) are the most popular for morphological pelvic sex estimation (Klales, 2013): ventral arc, subpubic concavity/contour, and the medial aspect of the ischio-pubic ramus. Because the Walker (2008) and Klales et al. (2012) methods and the sex classification results obtained when using them have been found to be highly correlated with the metrics of the skull and pelvis, Kenyhercz, Fredette, Klales, and Dirmaat (2012)

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00016-X

recommend the use of both method types (quantitative and qualitative) for sex estimation. Given the survey results mentioned above, it would appear that most practitioners are, in fact, using both types of methods, albeit perhaps with a preference for qualitative methods.

Quantitative sex estimation methods were consolidated into the computer program FORDISC in 1993 (Jantz & Ousley, 2005). Cranial and postcranial metric data from 13 modern populations (eight male groups and five female groups) have been integrated into the interactive program (Jantz & Ousley, 2005). Forensic scientists can enter the measurements of their unknown individual into the program and then compare their case to those individuals within the known reference samples for sex and ancestry classification using linear discriminant function analysis and stature estimation using linear regression (see Chapter 12 of this volume for more information). Users are then provided with discriminant function classification accuracy, a posterior probability of group membership, and several typicalities that the practitioner can then interpret for combined sex and ancestry estimation. Within two short decades, FORDISC has become the number-one method for metric assessment of sex and ancestry (Klales, 2013). The ease of use and inclusive nature of the program is likely why FORDISC is the number-one method for metric assessment of sex and ancestry parameters of the biological profile (Klales, 2013).

Unfortunately, a similar program for morphological methods did not yet exist; therefore, practitioners must rely on the equations provided in the original publications and then must evaluate sample and statistical appropriateness when applying the method to their unknown individual. To remedy this, the MorphoPASSE: Morphological Pelvis and Skull Sex Estimation program was created through a National Institute of Justice-funded grant (2015-DN-BX-K014). The primary aim of the database project was to examine temporal changes, population variation, and the effects of asymmetry on sex classification using the eight most popular morphological traits of the skull (five used in the Walker, 2008 method) and the pelvis (three used in the Klales et al., 2012 method). The secondary aim of this project was to develop MorphoPASSE, a free, interactive morphological program, based on these standardized methods. With the program, practitioners can enter, compare, and analyze morphological traits of their unknown individual to a large sample with known demographic data in order to more accurately and more easily estimate sex.

About MorphoPASSE

Skull and pelvis score data in the MorphoPASSE program come from 15 different collections (Table 1) and contain individuals from five broad geographic ancestral backgrounds: Asian ($n=266$), African/Black ($n=685$), Hispanic ($n=320$), Native American ($n=117$), and European/White ($n=1207$) (Table 2). Both contemporary and historical samples are included, thereby making the database applicable to modern

Table 1 Skeletal samples included in MorphoPASSE.

- Antioquia Modern Skeletal Reference Collection
- Arikara Collection at the University of Tennessee, Knoxville
- Hamann-Todd Human Osteological Collection
- Hartnett-Fulginiti Pubic Bone Collection
- Khon Kaen University Human Skeleton Research Centre
- Mercyhurst University Forensic Anthropology Laboratory
- Nubian Collection at the University of Colorado
- Osteological Collection of the National Autonomous University of Mexico
- Pretoria Bone Collection
- Robert J. Terry Anatomical Skeletal Collection
- Santa María Xigui Cemetery
- Texas State Operation Identification Collection
- Texas State University Donated Skeletal Collection
- University of the Philippines Skeletal Reference Collection
- William M. Bass Donated Skeletal Collection

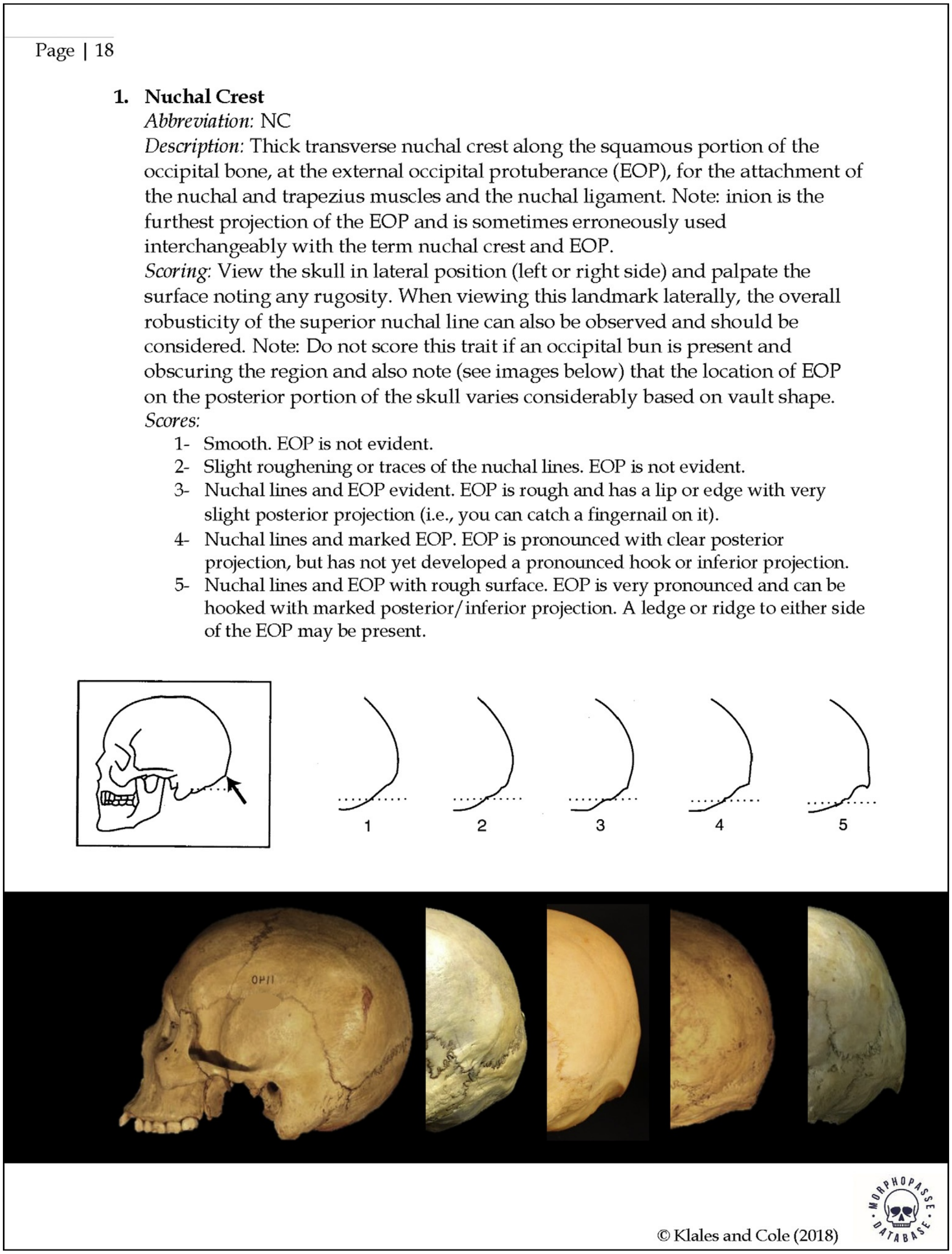
Table 2 Sample sizes in MorphoPASSE by geographic population/ancestry group.

	Males	Females	Total
Asian	179	87	266
Black	367	318	685
Hispanic	198	122	320
Native American	59	58	117
White	694	513	1207
Total sample			2595

forensic casework, as well as bioarchaeological cases. Prior to the formation of the program, population variation, temporal variation, observer error, and the impacts of asymmetry were assessed. Data not collected as part of the grant was sourced from fellow researchers after observer error studies demonstrated that all traits are reliable with the exception of the mental eminence (Walls & Klales, 2018). The MorphoPASSE program and accompanying manual can be found here: https://www.morphopasse.com/ and the database itself is accessible via R Studio's www.shinyapps.io/MorphoPASSE.

Scoring procedures

The five Walker (2008) and three Klales et al. (2012) traits should be scored using the MorphoPASSE manual (Klales & Cole, 2018), not the original publications, because modifications were made to the traits based on research from the grant. For example, Walker's (2008) traits were expanded to include descriptions for intermediate scores (2–4) and also revised to include real bone examples of the traits to accompany his original drawings (Fig. 1, example for the nuchal crest). The manual is freely available at www.MorphoPASSE.com.



Page | 18

1. Nuchal Crest

Abbreviation: NC

Description: Thick transverse nuchal crest along the squamous portion of the occipital bone, at the external occipital protuberance (EOP), for the attachment of the nuchal and trapezius muscles and the nuchal ligament. Note: inion is the furthest projection of the EOP and is sometimes erroneously used interchangeably with the term nuchal crest and EOP.

Scoring: View the skull in lateral position (left or right side) and palpate the surface noting any rugosity. When viewing this landmark laterally, the overall robusticity of the superior nuchal line can also be observed and should be considered. Note: Do not score this trait if an occipital bun is present and obscuring the region and also note (see images below) that the location of EOP on the posterior portion of the skull varies considerably based on vault shape.

Scores:

1- Smooth. EOP is not evident.
2- Slight roughening or traces of the nuchal lines. EOP is not evident.
3- Nuchal lines and EOP evident. EOP is rough and has a lip or edge with very slight posterior projection (i.e., you can catch a fingernail on it).
4- Nuchal lines and marked EOP. EOP is pronounced with clear posterior projection, but has not yet developed a pronounced hook or inferior projection.
5- Nuchal lines and EOP with rough surface. EOP is very pronounced and can be hooked with marked posterior/inferior projection. A ledge or ridge to either side of the EOP may be present.

© Klales and Cole (2018)

Fig. 1 Page 18 from the MorphoPASSE manual (Klales & Cole, 2018) showing modifications to Walker's (2008) nuchal crest trait.

The manual includes a description of the trait, scoring procedures, and special considerations followed by individual score descriptions from the original publications, modifications and revisions, schematic representations of each trait, and real bone specimen photos of each trait. Prior to scoring the traits, the analyst should become familiar with the range of variation present by minimally examining the real bone specimens provided in the manual. For each trait, the analyst should view the specimen and compare it to both the descriptions and figures (drawings and real bone examples) to score the specimen. For some traits, multiple features are being scored; therefore, weight or preference should be given to those listed as most important in the manual. For example, the mental eminence examines the tubercles, as well as the portion of the mandible occupied by the eminence. Likewise, the ventral arc examines the ridge of bone, as well as the overall bone shape and morphology. In the case of bilateral traits, both the left and right sides should be scored. Lastly, the manual includes a scoring form (page 29) and information on how the data can be accessed by outside researchers for additional projects.

Statistical options

In keeping consistent with Walker (2008) and Klales et al. (2012), MorphoPASSE allows the analyst to select the binary logistic regression (LR) equations provided in the original publications for sex classification and provides calculation of posterior probabilities of sex membership. In both these methods, the sex of the individual is treated as the binary dependent variable, and the skeletal indications are the ordinal independent variables. However, as Konigsberg and Frakenberg (2019: 385) recently pointed out, skeletal "indicators depend on the sex of the individual, rather than the sex of the individual depending on the indicators"; therefore, a different statistical approach may be more appropriate for these traits. Nonetheless, LR was chosen in these articles due to its numerous advantages over other classification methods (e.g., discriminant function analysis). The relaxed assumptions of LR do not require normally distributed data, and it remains robust despite normality deviations. LR does generally assume larger samples and that (1) variables are discrete, (2) there are no data outliers, (3) a linear relationship exists between each independent variable and the odds ratio, and (4) there is to be no collinearity among predictor variables. While not all of these assumptions are met, Walker (2008) and Klales et al. (2012) argued that practical criteria are more important than dogmatically adhering to the rules (see Chapter 13 of this volume for a more in-depth discussion of this Measurement-Statistic debate).

Because of the collinearity of these 13 variables and the inability of LR to easily handle missing data, MorphoPASSE also includes random forest modeling (RFM). RFM is the recommended application in MorphoPASSE and is a flexible machine learning (ML) algorithm that creates a series of decision trees using bootstrap aggregating of random training subsets and then produces an average prediction based on the "forest." Random forest classification uses many random subsets of the variables and repeated sampling of the original data to produce hundreds of decision trees, called an ensemble, and the consensus of the ensemble is used to determine the best classification rules. Random forests can generally tolerate a large number of variables simultaneously, including "noisy" ones (Hefner & Ousley, 2014: 886). Thousands of random cut-off points in the sample are determined "on-the-fly" to determine the most accurate pooling of groups (i.e., the sexes in this case) (Hefner & Ousley, 2014; Williams, 2011). The more trees in the forest, the more robust or accurate the sex prediction. This approach prevents overfitting and only selects the most valuable input features, or traits and their scores, for classification. RFM is non-parametric whereby the model is based on the data entered (i.e., not specified a priori) and makes no assumptions about that data (e.g., requirement of normal distribution, sample size, etc.). Thus far, ML approaches, including decision trees/random forest models, have been mostly applied to continuous data for sex estimation; however, these statistical approaches have also shown great promise for morphological traits (binary, discrete, ordinal data) and combined morphological/metric ancestry estimation (Hefner & Ousley, 2014; Hefner, Spradley, & Anderson, 2014), but have yet to be widely applied in this capacity to sex estimation (see Chapter 13 of this volume).

Users may enter in any of the 13 variables (three unilateral and two bilateral skull traits and three bilateral pelvis traits) into MorphoPASSE to generate an "on-the-fly" prediction of sex membership based on the population and temporal-specific criteria selected by the practitioner. Posterior probabilities, determining how likely the entered individual's score are to belong to each sex, are also provided to interpret the strength of the results. For example, a probability of 60% (i.e., close to random chance) should be interpreted as far less meaningful than a probability of membership of 85% or above for sex. Practitioners also have the option of utilizing the LR equations provided in the original Walker (2008) and Klales et al. (2012) articles rather than using the "on-the-fly" calculations.

Interface

On the input page, the analyst enters their name (or initials) and case identification number. Next, they select the statistical option to use—again RFM is the recommended approach. Then a temporal period, ancestry group, and/or region can be selected based on the case being analyzed. If none of these are selected (i.e., all listed as unknown), the program will use the entirety of the database sample. Lastly, at least one trait score must be entered for analysis. Once trait scores are entered, the total sample size used in the analysis will be displayed. The final step for classification is selecting the "run analysis" button at the top. Once the analysis is complete, the output page will come up automatically. A word document of the analysis and report can be downloaded.

Included in the output are the following: data entered, test parameters selected, model formula, case prediction/sex probability, test accuracy, training model accuracy, and variable importance. *Case Prediction* provides the probability of sex membership. *Model* is the model summary. The type of RFM is classification, because sex is a binary variable. The number of trees is included along with the number of predictor variables considered at each node of the decision tree. The "*out of the bag*" (OOB) estimate of the error is based on bootstrap aggregation. At each iteration created with a subset of data, the unused data is tested in the tree to produce an average of errors for the entire set of decision tree. Model tuning *mtry* is the number of variables randomly sampled as candidates for each node and is also presented visually. The confusion matrix presents the accuracy of the model based on true negatives, true positives, false positives, and false negatives. *Variable Importance* (mean decrease in Gini coefficient) describes how important each of the variables is when classifying sex. The most important variable will be the one with the highest mean decrease in OOB error. Typically, pelvic traits will always be of more importance than skull traits due to the higher degree of sexual dimorphism in the pelvis. This information is also presented visually. *Model Training* provides cross-validated classification accuracy of the entire sample. The *Kappa* statistic provides the accuracy of the model taking into account random chance and will typically be lower than the accuracy. The details of model training, percent accuracy, and kappa statistic are only provided in the downloaded report. *Model Accuracy* tests the model on a hold-out sample from the database. The following are also provided: sensitivity/true positive, (TP)/(TP + FN); specificity/true negative, (TN)/(FP + TN); positive predictive value, (TP)/(TP + FP); and negative predictive value, (TN)/(FN + TN).

Conclusion

The MorphoPASSE program provides a free, user-friendly means by which to utilize the Walker (2008) and Klales et al. (2012) traits and associated methods for reliable and valid sex estimation. In the future, the hope is that (1) additional morphological traits of the skull and pelvis, for example, the greater sciatic notch (Walker, 2005), will be added to the program, and (2) data can be sourced from additional worldwide populations to increase the sample size and global representativeness of the database. Moving toward databases and programs, like FORDISC and MorphoPASSE, will aid in the quest for standardization of our methods.

Acknowledgments

This research was funded by the National Institute of Justice Grant 2015-DN-BX-K014. Opinions or points of view expressed in this research represent a consensus of the authors and do not necessarily represent the official position or policies of the US Department of Justice or the National Institute of Justice. Any products and manufacturers discussed are presented for informational purposes only and do not constitute product approval or endorsement.

References

Buikstra, J. E., & Ubelaker, D. H. (1994). Standards for data collection from human skeletal remains. In *Proceedings of a seminar at the Field Museum of natural history*.

Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993). *509 U.S. 579*.

Hefner, J. T., & Ousley, S. D. (2014). Statistical classification methods for estimating ancestry using morphoscopic traits. *Journal of Forensic Sciences*, *59*(4), 883–890.

Hefner, J. T., Spradley, M. K., & Anderson, B. (2014). Ancestry assessment using random forest modeling. *Journal of Forensic Sciences*, *59*(3), 583–589.

Jantz, R. L., & Ousley, S. D. (2005). *FORDISC 3: Computerized forensic discriminant functions*. Version 3.0 Knoxville, TN: The University of Tennessee-Knoxville.

Kenyhercz, M. W., Fredette, S. M., Klales, A. R., & Dirmaat, D. C. (2012). Metric and non-metric assessment of sex: Accuracy, correlation and corroboration. [Abstract]. *Proceedings of the American Academy of Forensic Sciences*, *18*, 370.

Klales, A. R. (2013). Current practices in forensic anthropology for sex estimation in unidentified, adult individuals. [Abstract]. In *19. Proceedings American Academy Forensic Sciences*.

Klales, A. R., & Cole, S. J. (2018). *MorphoPASSE: The morphological pelvis and skull sex estimation database manual*. Version 1.0 Topeka, KS: Washburn University.

Klales, A. R., Ousley, S. D., & Vollner, J. M. (2012). A revised method of sexing the human innominate using phenice's nonmetric traits and statistical methods. *American Journal of Physical Anthropology*, *149*(1), 104–114.

Konigsberg, L. W., & Frakenberg, S. R. (2019). Multivariate ordinal probit analysis in the skeletal assessment of sex. *American Journal of Physical Anthropology*, *169*(2), 385–387.

Phenice, T. W. (1969). A newly developed visual method for sexing the os pubis. *American Journal of Physical Anthropology*, *30*, 297–302.

Rogers, T. L., & Saunders, S. (1994). Accuracy of sex determination using morphological traits of the human pelvis. *Journal of Forensic Sciences*, *39*, 1047–1056.

Walker, P. L. (2005). Greater sciatic notch morphology: Sex, age, and population differences. *American Journal of Physical Anthropology*, *127*, 385–391.

Walker, P. L. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, *136*, 39–50.

Walls, M. M., & Klales, A. R. (2018). Testing the reliability of morphological sex estimation methods used in forensic anthropology. [Abstract]. In *24. Proceedings American Academy Forensic Sciences*.

Williams, G. (2011). *Data mining with rattle and R*. New York, NY: Springer.

Williams, B. A., & Rogers, T. (2006). Evaluating the accuracy and precision of cranial morphological traits for sex determination. *Journal of Forensic Sciences*, *51*(4), 729–735.

CHAPTER 17

Factors of population variation in sex estimation methodology

Douglas H. Ubelaker, Cassandra M. DeGaglia
Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC, United States

The estimation of biological sex from anthropological analysis of human skeletal remains represents a key element in the process leading to positive identification. An accurate estimation of sex allows the investigation to focus on that segment of the list of missing persons. Sex evaluation also enables hypotheses to be tested regarding possible or presumed identifications. For these reasons, historically, sex estimation has been regarded as an important element in the biological profile established through anthropological analysis.

The methodology of sex estimation involves metric approaches, visual observations, or some combination of both, although molecular methods are now also available (Gaballah, Shehab, & Bayoumi, 2014). The historical foundation of most of the methods is rooted primarily in 20th-century North America with the availability of documented skeletal collections and researchers interested in this topic. The pioneering research of Thomas Dwight (1843–1911), Wilton Krogman (1903–1988), T.D. Stewart (1901–1997), and others provided initial data and methodology on the sexual dimorphism of the human skeleton (Ubelaker, 2009). Documented skeletal assemblages, such as the Terry collection curated in Washington, DC, and the Hamman-Todd collection in Cleveland, Ohio, provided the opportunity for this important research. By the time Stewart published his classic text *Essentials of Forensic Anthropology: Especially as Developed in the United States* in 1979, a variety of metric and observation-oriented methods were available.

More recently, forensic anthropologists have recognized the limitations and regionally specific nature of this early methodology. While methods published from North American samples represented significant aspects of sexual dimorphism in that region, how well did they reflect global variation? This issue became increasingly important as skeletal analysis and identification has not been limited to North America. When practitioners around the world attempted to apply these methods, uncertainty developed regarding the accuracy and impact of regional variation. Concerns were exacerbated by the growing literature in growth and development strongly indicating that the expression of sexual dimorphism in human morphology is influenced by both genetics and environmental factors. The latter include diet, disease, and socioeconomic status that

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00017-1

are known to vary greatly throughout the world. This chapter aims to provide a sample of these recent global efforts to document population-specific skeletal data.

Regional variation of human growth patterns

Research in many countries has documented the powerful role of environment on the expression of sexual dimorphism (Eleveth & Tanner, 1990). During the growth process, the mean height of girls is generally greater than that of boys prior to the adolescent growth spurt in the latter. The more advanced growth of girls produces greater stature until the delayed growth of boys. At the end of adolescence, boys universally are taller than girls but with great global variation in the actual values. Comparative data reveal secular trends as well as regional patterns reflecting largely differing environmental conditions (Eleveth & Tanner, 1990; Silventoinen, 2003; Ubelaker & DeGaglia, 2017). These secular trends and regional diversity relate not only to general body dimensions but to the skeletal system as well (Ubelaker & DeGaglia, 2017).

Skeletal sexual dimorphism

Developments discussed above reveal the need for regionally based methods and databases regarding skeletal sexual dimorphism. Such new initiatives now are possible due to the establishment of diverse collections, innovative technology, and the global surge of interest in forensic anthropology. While North American collections of documented human remains continue to grow and enjoy sustained scholarly attention, similar research resources are now available in many countries (Ubelaker, 2014). These new collections of human remains of known sex and other demographic variables have enabled remarkable research revealing considerable detail on the regional variation of skeletal sexual dimorphism.

In addition to these valuable collections, new technology facilitates not only skeletal analysis but studies of the living as well. Advances include three-dimensional morphological analysis of skeletal tissue, sophisticated statistics, and augmented computerized databases that supplement traditional measurement and observation. Clinical imagery, such as CAT scan approaches, enables a detailed analysis of skeletal anatomy in the living. The strong global interest in forensic anthropology attracts creative and intelligent young professionals resulting in innovative research designs.

The following text summarizes many of the recent publications that have resulted from this new effort. While this summary is not comprehensive, it reflects the diversity and global nature of sustained research documenting variation in skeletal dimorphism. Organization of this information features the skeletal element investigated, beginning with the cranium and culminating with the pelvis and anatomically comprehensive approaches.

Skull

Working in South Africa in 2005, Franklin, Freedman, and Milne (2005) examined sexual dimorphism and discriminant function sex estimation methods. Their craniometric study focused on skeletal remains of indigenous peoples of that country. Using pooled measurements, they found that sex could be estimated accurately in 77%–80% of the skeletons examined.

In 2006, Kemkes and Göbel (2006) tested a previously published method involving metric assessment of the mastoid triangle. They studied 25 female and 72 male skulls from Germany, and 50 males and 50 females from Portugal. The study found that although sex differences were detected, they were not of the magnitude useful in forensic investigation. Their interpretation focused on population-specific variability in this anatomical feature.

The frontal sinus has received considerable attention in forensic anthropology, primarily for its great morphological variability that facilitates positive identification. In a Brazilian study, Camargo et al. (2007) examined the usefulness of this feature for sex estimation. Their examination of 50 males and 50 females of known sex demonstrated that mean measurements of males were greater than those of females. Belaldavar, Kotrashetti, Hallikerimath, and Kale (2014) later found similar results in their study of 150 males and 150 females in India.

In 2008, Kranioti, Iscan, and Michalodimitrakis (2008) examined sexual dimorphism in Cretan crania. Employing a battery of 16 craniofacial measurements, they found that bizygomatic breadth allowed a sex estimation accuracy of 82%. Accuracy increased to 88.2% when a stepwise procedure utilizing five measurements was employed.

Dayal, Spocter, and Bidmos (2008) studied 60 males and 60 females from a South African Black (Dart) collection. Using discriminant function analysis, they found that 14 cranial and 6 mandibular measurements allowed sex to be estimated with an accuracy of 85%.

Working with a Western Australian sample, Franklin, Cardini, Flavel, and Kuliukas (2013) recorded 18 measurements in 200 male and 200 female crania. They found an accuracy of sex estimation as high as 90% using bizygomatic breadth and maximum length.

In India, Sharma, Gorea, Gorea, and Abuderman (2016) examined sex differences in measurements of the mandible. They provided data and functions for various measurements working with samples from Punjab and Chandigarh.

Hyoid

The forensic significance of the hyoid primarily relates to its tendency to fracture during manual strangulation and other forms of neck trauma (Ubelaker, 1992). Kim et al. (2006)

demonstrated that the bone also may reveal evidence of biological sex. This Korean study examined digital photographs of hyoids from 52 males and 33 females. Twenty-one of thirty-four measurements revealed sex differences. They developed a discriminant function that would estimate sex with an accuracy of 88.2%.

Scapula/Clavicle

In 2012, Papaioannou, Kranioti, Joveneaux, Nathena, and Michalodimitrakis (2012) examined sexual dimorphism of the scapula and clavicle in a contemporary Greek sample. They found that in their sample of 147 left scapulae and 147 left clavicles (81 males and 66 females), sex could be estimated correctly from maximum scapular height with an accuracy of 91.2%. The use of two variables increased accuracy to 95.9%.

In an Egyptian study, Badr El Dine and Hassan (2016) employed multidetector CT to examine sex variation of the scapula. Comparative data were generated from a sample of 83 males and 79 females.

Working in Japan, Torimitsu et al. (2016a) studied scapular sexual dimorphism in a sample of 109 males and 109 females. A stepwise procedure enabled sex to be estimated with an accuracy of 94.5%. In a similar study of the clavicle using 150 males and 150 females, they were able to estimate sex correctly in 92.2% of individuals (Torimitsu et al., 2018).

Sternum

In 2010, Osunwoke et al. evaluated sexual dimorphism in a sample of 94 sterna (68 males and 26 females) obtained from five universities in southern Nigeria. They found the mean lengths of manubria to be 60.7 mm in males and 46.0 mm in females. The combined mean lengths of the manubrium and sternal body (not including the xiphoid process) were 164.6 mm in males and 123.3 mm in females (Osunwoke, Gwunireama, Orish, Ordu, & Ebowe, 2010).

In 2012, Franklin et al. used MSCT scans to estimate sex in a sample of 93 male and 94 female sterna from western Australia. Mean manubrium length was 49.02 mm in males and 45.32 mm in females. Combined manubrium and body length (again excluding the xiphoid process) revealed a mean of 151.96 mm in males and 130.22 mm in females. When employing eight linear measurements of the sternum and a stepwise discriminant function analysis, the authors were able to obtain a classification accuracy of 84.5%. However, there was a sex bias of −3.4% (Franklin et al., 2012).

Also, in 2012, Singh et al. conducted morphometric sex estimation from sternal widths of a northwestern Indian autopsy sample. In their sample of 343 individuals (252 males and 91 females), regression analysis allowed sex estimation of 86.6% accuracy (Singh, Pathak, & Sing, 2012).

Vertebrae

Although the general size of the vertebrae is known to be sexually dimorphic, relatively few detailed studies have been conducted. Working with the 12th thoracic and first lumbar vertebrae, Badr El Dine and El Shafei (2015) used multislice CT to capture measurements in an Egyptian sample of 54 males and 66 females. Fourteen measurements revealed sex differences. They were able to estimate sex correctly in 93.1% of their sample from T12 and 68% using L1. Combined measurements from the two bones produced an accuracy of 96.3%.

In Japan, Torimitsu et al. (2016b) employed multidetector CT to measure the second cervical vertebra. Measurements of their sample of 112 males and 112 females allowed sex to be estimated with an accuracy of 92.9%.

Humerus, radius, and ulna

In 1998, İşcan, Loth, King, Shihai, and Yoshino (1998) examined sexual dimorphism of the humerus in a comparative study of Chinese, Japanese, and Thai samples. Their stepwise procedure allowed sex to be estimated correctly in 86.8% of the Chinese sample, 92.4% of the Japanese sample, and 97.1% of Thai humeri.

In 1999, Steyn and İşcan (1999) conducted a similar study with South African samples. They were able to estimate sex correctly in 96% of the white sample and 95% of the black component. They found that in the white sample, measurements of the head and epicondylar breadth provided the best sexual indicator. In the black sample, the most accurate indicators originated from measurements of the head and maximum length.

Working in Guatemala, Frutos (2005) added comparative data from six measurements of 68 males and 50 females. The most accurate (95.5%) individual indicator was maximum head diameter. The use of a stepwise procedure increased accuracy to 98.2%.

Kranioti and Michalodimitrakis (2009) provided a literature review and new data from a contemporary Cretan sample (Greece) of 84 males and 84 females. They found the vertical head diameter to be the most accurate individual indicator (89.9%). The use of all measurements produced an accuracy of 92%. Also, in 2009, Kranioti, Bastir, Sánchez-Meseguer, and Rosas (2009) reported similar results and provided a comparative discussion. A 2011 report by Kranioti, Nathena, and Michalodimitrakis (2011) using a digital radiometric technique on a sample of 53 males and 48 females produced an accuracy of 89.1%.

Ali and Elbaky (2016) noted that sex estimation is sometimes needed for fragments of the humerus. Their study of 75 males and 75 females from Egypt demonstrated that the most accurate segment, from below the major tubercle to the upper margin of the olecranon fossa, allowed sex to be estimated accurately in 86.7% of their sample. Total length of the humerus produced an accuracy of 93.3%.

Also, in 2016, Issa, Khanfour, and Kharoshah (2016) provided data for the radius and ulna in an Egyptian autopsy sample. Measurements of bone lengths in a sample of 85 males and 37 females produced a combined accuracy of 98%.

Femur and tibia

In 1995, İşcan and Shihai documented sexual dimorphism in the Chinese femur (Tables 1 and 2). Their study generated data for six measurements of both males and females. They found that distal epiphyseal breadth produced the most accurate estimate of sex (94.9%).

Steyn and İşcan (1997) later added data from South Africa. In their study of the femur and tibia of 56 males and 50 females, they also found that the distal breadth of both bones produced the most accurate estimates. A combined method revealed 91% accuracy.

Femoral head diameter data from Nigeria emerged through the Asala et al. (1998) study of 257 males and 247 females. The values reported added key comparative data from an area of the world with little previous study. In 2000, Igbigbi and Msamati added similar data from Malawi. Asala (2001) compared data between white and black South Africans. Purkait and Chandra (2002) contributed femoral data from India.

Table 1 Mean vertical femoral head diameters (mm) by population sample.

	Male	Female
Malawi (black)	48.30	44.50
Northeast Nigeria	54.08	46.85
South Africa (black)	44.45	39.64
South Africa (white)	48.40	42.28
China	46.16	41.13
North India	43.77	39.40
Austria	47.50	42.00
Netherlands	49.06	43.25

Igbigbi and Msamati (2000), Asala, Mbajiorgu, and Papandro (1998), Asala (2001), İşcan and Shihai (1995), Srivastava, Saini, Rai, Pandey, and Tripahi (2012), Kanz, Fitzl, Vlcek, and Frommlet (2015), and Colman et al. (2018).

Table 2 Mean distal femoral epiphyseal breadth (mm) by population sample.

	Male	Female
South Africa (white)	51.2	44.36
China	80.32	70.62
India	76.83	68.28
India	76.27	69.26
Austria	81.1	72.8
France	84.3	74.8

Steyn and İşcan (1997), İşcan and Shihai (1995), Srivastava et al. (2012), Soni, Dhall, and Chhabra (2010), Kanz et al. (2015), and Alunni-Perret, Staccini, and Quatrehomme (2008).

In 2008, Alunni-Perret et al. examined the distal femur in a French sample building on previous reports of the value of this anatomical area for sex estimation. They found 95.4% accuracy in their sample.

Robinson and Bidmos (2008) examined the accuracy of discriminant function equations for sex estimation of both the femur and tibia in South African samples. They studied 272 femora and 256 tibiae representing four collections. They reported that four discriminant functions for the femur and one for tibia performed well in estimating sex.

Jantz, Kimmerle, and Baraybar (2008) published new femoral data from the Balkan area. Their study of samples from Kosovo and Bosnia revealed that femoral head diameters were larger than found in American whites for both males and females. New femoral data were also reported for northeast India (Soni et al., 2010) and North India (Srivastava et al., 2012).

Meeusen, Christensen, and Hefner (2015) examined femoral neck axis length to assess sexual dimorphism. Their American study revealed differences among samples of blacks, whites, and Native Americans.

Femoral data from Austria emerged from the Kanz et al. (2015) study of 72 females and 55 males. They did not find significant secular change and reported that maximum diameter of the femoral head provided the most accurate sex estimates at 87.8%.

Colman et al. (2018) added information on the proximal femur from the Netherlands. They utilized CT scan data in their study of 57 males and 57 females. Measurements of the femoral head allowed sex to be estimated with an accuracy of 86%.

Collectively, this international research supports the study of Kotěrová et al. (2017) of the tibia in indicating the importance of population-specific studies and data.

Patella

Considerable comparative international data have emerged in recent years regarding sex estimation from the patella. Working in southern Italy, Introna Jr., Di Vella, and Campobasso (1998) studied 40 males and 40 females. They found that maximum width and thickness of the patella allowed a sex estimation accuracy of 83.8%. In a German archaeological sample of individuals with estimated sex, Kemkes-Grottenhaler (2005) conducted a reliability study with favorable results.

Bidmos, Steinberg, and Kuykendall (2005) studied patella measurements of South African whites. Samples of 60 males and 60 females from the Dart collection revealed that maximum height allowed sex to be estimated with 85% accuracy.

Mahfouz et al. (2007) used 3D statistical shape models and non-linear classifiers to examine sexual dimorphism of the patella in the US Bass collection. Their application of this approach to 95 females and 133 males allowed sex to be estimated with 90.3% accuracy.

Data emerged from Iran with a study by Akhlaghi, Sheikhazadi, Naghsh, and Dorvashi (2010) of 57 males and 56 females. Their stepwise procedure produced estimates with an accuracy of 92.9%.

In Japan, Michiue et al. (2018) studied sexual dimorphism of the patella using virtual CT. Their study of 110 males and 110 females revealed estimates using 2D variables of 82.3%. Accuracy increased to 87.7% with the evaluation of bone mass.

Pelvis

Many publications indicate that the adult pelvis provides the most accurate information regarding sexual dimorphism, relating to the childbirth function of females (Rowbothan, 2016; Stewart, 1979; Ubelaker, 1999). Informative variables are primarily observational but can include metric size-related factors as well. Approaches that reflect size and robusticity would be expected to display some population variation. Methods of estimating sex from the pelvis using observations of anatomical details related to the female childbirth process would be expected to show less variation. However, even these observational features present some variation. For example, in 1969, Phenice published a method utilizing three features of the pubic area. By examining the ventral arc, subpubic concavity, and medial aspect of the ischio-pubic ramus on 275 individuals of European and African descent from the Terry Collection, Phenice was able to estimate sex from adult pelves with 96% accuracy (Phenice, 1969). Later, Ubelaker and Volk (2002) experimentally revealed that experience is an important factor even in this approach. Lovell (1989) also tested the method on 13 males and 23 females, finding an accuracy of sex estimation of only 83%. In 2012, Klales et al. presented a revised method of using Phenice's traits that included statistical analysis. The authors built upon Phenice's binary observations by assigning five character states with ordinal scores. This scoring method allowed the reliability and accuracy of the results to be calculated (Klales, Ousley, & Vollner, 2012). Subsequently, Kenyhercz, Klales, Stull, McCormick, and Cole (2017) noted the impact of population variation on sex estimation from the pelvis. Using the Klales et al. (2012) method, sex was estimated for a large global sample of 1915 individuals of American black and white, South African black and white, Thai, and Hispanic ancestries. The authors noted that the black and white populations from both the United States and South Africa produced the highest sex classification accuracies, while the Thai and Hispanic populations presented the lowest sex classification accuracies (Kenyhercz et al., 2017).

In a 1994 Canadian study of 49 individuals, Rogers & Saunders, 1994 reported a 94.1% estimation accuracy of sacrum shape with other pelvic features being less accurate. In their study, multiple indicators produced an accuracy of 88%; and all features, 95.9%.

Using identified adult skeletons, Gómez-Valdés et al. (2017) recalibrated the Klales et al. (2012) method for contemporary Mexican populations. Although applying the Klales et al. method without modification to their sample presented 100% accuracy for females and only 86%–92% accuracy for males, the recalibration increased correct classification of both sexes to 100% when examining all three of Phenice's traits (Gómez-Valdés et al., 2017).

Bruzek (2002) reported his study of 402 adults from collections in France and Portugal. The use of five features combined enabled accuracy close to 98%. The five features were preauricular surface, greater sciatic notch, composite arch (anterior auricular area and sciatic notch), inferior margin of the innominate, and ischiopubic proportions.

In 2008, Papaloucas, Fiska, and Demetriou (2008) reported a Greek study of 100 males and 100 females. They found that consideration of the femoral head diameter and the acetabulum diameter allowed an accuracy approaching 99%. Also, in 2008, Steyn & İşcan, 2008 reported metric pelvic data from modern Greeks. Their sample of 97 males and 95 females revealed that acetabulum measurements enabled an accuracy of 83.9%.

The Japanese study by Torimitsu et al. (2015) involving morphometric analyses of 104 males and 104 females using multidetector CT found that among the individual variables the subpubic angle offered the most useful data. Their multivariate approach produced an accuracy of 98.1%.

General approaches

In 2005, Murail, Bruzek, Houët, and Cunha (2005) introduced the approach, "Diagnose Sexuelle Probabiliste" (DSP), that utilized worldwide data. This method is based on data from 20,140 individuals from 12 different reference samples originating in Europe, Africa, North America, and Asia. Through a comparison with this broadly constructed database, the method calculates the probability of the unknown sample. They report accuracy close to 100%. In 2006, Bruzek & Murail, 2006 offered general summary recommendations for sex estimation and noted the superior results of estimates using pelvic indicators (see Chapter 15 of this volume).

Discussion

The sampling of recent international research on skeletal sexual dimorphism, presented above, documents the emergence of new collections and focused research in many countries. The data presented make the case that patterns of sexual dimorphism vary regionally, likely in response to environmental factors along with genetic variation. Clearly, standards and methodology developed in one region from one collection are not universally applicable. Tables 1 and 2 present global variation in mean values of two dimensions commonly used to estimate sex: femoral vertical head diameter and femoral distal epiphyseal breadth. These data document aspects of global variation and demonstrate why local databases are so important in forensic anthropological analysis.

Some issues of regional variation are addressed through formation of methods based on multi-region datasets. While these approaches appear superior to applying methods developed from a very regionally restricted sample to unknowns from very different regions, issues remain. Practitioners should strive to utilize methods that are most appropriate to the cases they are presented with. There is no single answer to this challenge. Education is the key—especially awareness of the strengths and weaknesses of the methods available.

Most global research on this issue reveals that use of multiple measurements and/or indicators enables more accurate estimates than single ones. Comprehensive approaches are usually more powerful throughout forensic science, and the estimation of sex is no exception.

Caution is required to interpret much of the accuracy results reported in the literature. High accuracy reported in a particular regional study may accurately describe the results in that project, but the method would likely be less accurate in applications to unknowns from different regions. Also, it is difficult to evaluate the impact of experience with a particular method or nuance in the equipment utilized. The latter problem is especially acute with advanced technology that requires considerable training and accurate calibration.

Conclusion

The need for global studies of sexual dimorphism of the human skeleton is being addressed. This research is possible now due to the formation of large, well-documented collections of human remains in many areas of the world. Advances have been stimulated by the increasing availability of statistical programs, computers, imaging equipment, and related facilities. Perhaps most importantly, global interest in forensic anthropological research is strong and growing. This interest can be measured by the growing numbers of students being attracted to this academic area, the formation and popularity of regional organizations devoted to forensic anthropology, and the published research revealed in our journals. We have learned that sexual dimorphism of the human skeleton does vary regionally. We are learning details about that variation and how it can be addressed in casework.

References

Akhlaghi, M., Sheikhazadi, A., Naghsh, A., & Dorvashi, G. (2010). Identification of sex in Iranian population using patella dimensions. *Journal of Forensic and Legal Medicine*, *17*(3), 150–155.

Ali, D. M., & Elbaky, F. A. F. A. (2016). Sex identification and reconstruction of length of humerus from its fragments: An Egyptian study. *Egyptian Journal of Forensic Sciences*, *6*(2), 48–55.

Alunni-Perret, V., Staccini, P., & Quatrehomme, G. (2008). Sex determination from the distal part of the femur in a French contemporary population. *Forensic Science International*, *175*(2–3), 113–117.

Asala, S. A. (2001). Sex determination from the head of the femur of South African whites and blacks. *Forensic Science International*, *117*(1-2), 15–22.

Asala, S. A., Mbajiorgu, F. E., & Papandro, B. A. (1998). A comparative study of femoral head diameters and sex differentiation in Nigerians. *Cells Tissues Organs*, *162*(4), 232–237.

Badr El Dine, F. M. M., & El Shafei, M. M. (2015). Sex determination using anthropometric measurements from multi-slice computed tomography of the 12th thoracic and the first lumbar vertebrae among adult Egyptians. *Egyptian Journal of Forensic Sciences*, *5*(3), 82–89.

Badr El Dine, F. M. M., & Hassan, H. H. M. (2016). Ontogenic study of the scapula among some Egyptians: Forensic implications in age and sex estimation using Multidetector Computed Tomography. *Egyptian Journal of Forensic Sciences*, *6*(2), 56–77.

Belaldavar, C., Kotrashetti, V. S., Hallikerimath, S. R., & Kale, A. D. (2014). Assessment of frontal sinus dimensions to determine sexual dimorphism among Indian adults. *Journal of Forensic Dental Sciences*, *6*(1), 25–30.

Bidmos, M. A., Steinberg, N., & Kuykendall, K. L. (2005). Patella measurements of South African whites as sex assessors. *HOMO—Journal of Comparative Human Biology*, *56*(1), 69–74.

Bruzek, J. (2002). A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology*, *117*(2), 157–168.

Bruzek, J., & Murail, P. (2006). Methodology and reliability of sex determination from the skeleton. In *Forensic anthropology and medicine* (pp. 225–242): Humana Press.

Camargo, J. R., Daruge, E., Prado, F. B., Caria, P. H. F., Alves, M. C., Silva, R. F., et al. (2007). The frontal sinus morphology in radiographs of Brazilian subjects: Its forensic importance. *Brazilian Journal of Morphological Sciences*, *24*(4), 239–243.

Colman, K. L., Janssen, M. C. L., Stull, K. E., van Rijn, R. R., Oostra, R. J., de Boer, H. H., et al. (2018). Dutch population specific sex estimation formulae using the proximal femur. *Forensic Science International*, *286*, 268.e1–268.e8.

Dayal, M. R., Spocter, M. A., & Bidmos, M. A. (2008). An assessment of sex using the skull of black South Africans by discriminant function analysis. *HOMO—Journal of Comparative Human Biology*, *59*(3), 209–221.

Eleveth, P. B., & Tanner, J. M. (1990). *Worldwide variation in human growth* (2nd ed.). Cambridge: Cambridge University Press.

Franklin, D., Cardini, A., Flavel, A., & Kuliukas, A. (2013). Estimation of sex from cranial measurements in a Western Australian population. *Forensic Science International*, *229*(1–3). 158.e1–158.e8.

Franklin, D., Flavel, A., Kuliukas, A., Cardini, A., Marks, M. K., Oxnard, C., et al. (2012). Estimation of sex from sternal measurements in a Western Australian population. *Forensic Science International*, *217*(1–3). 230-e1.

Franklin, D., Freedman, L., & Milne, N. (2005). Sexual dimorphism and discriminant function sexing in indigenous South African crania. *HOMO—Journal of Comparative Human Biology*, *55*(3), 213–228.

Frutos, L. R. (2005). Metric determination of sex from the humerus in a Guatemalan forensic sample. *Forensic Science International*, *147*(2–3), 153–157.

Gaballah, I. F., Shehab, A. M., & Bayoumi, K. A. (2014). Sex determination in femurs of modern Egyptians: A comparative study between metric measurements and SRY gene detection. *Egyptian Journal of Forensic Sciences*, *4*(4), 109–115.

Gómez-Valdés, J. A., Menéndez Garmendia, A., García-Barzola, L., Sánchez-Mejorada, G., Karam, C., Baraybar, J. P., et al. (2017). Recalibration of the Klales et al. (2012) method of sexing the human innominate for Mexican populations. *American Journal of Physical Anthropology*, *162*(3), 600–604.

Igbigbi, P. S., & Msamati, B. C. (2000). Sex determination from femoral head diameters in black Malawians. *East African Medical Journal*, *77*(3).

Introna, F., Jr., Di Vella, G., & Campobasso, C. P. (1998). Sex determination by discriminant analysis of patella measurements. *Forensic Science International*, *95*(1), 39–45.

İşcan, M. Y., Loth, S. R., King, C. A., Shihai, D., & Yoshino, M. (1998). Sexual dimorphism in the humerus: A comparative analysis of Chinese, Japanese, and Thais. *Forensic Science International*, *98*(1–2), 17–29.

İşcan, M. Y., & Shihai, D. (1995). Sexual dimorphism in the Chinese femur. *Forensic Science International*, *74*(1), 79–87.

Issa, S. Y., Khanfour, A. A., & Kharoshah, M. (2016). A model for stature estimation and sex prediction using percutaneous ulnar and radial lengths in autopsied adult Egyptians. *Egyptian Journal of Forensic Sciences*, *6*(2), 84–89.

Jantz, R. L., Kimmerle, E. H., & Baraybar, J. P. (2008). Sexing and stature estimation criteria for Balkan populations. *Journal of Forensic Sciences*, *53*(3), 601–605.

Kanz, F., Fitzl, C., Vlcek, A., & Frommlet, F. (2015). Sex estimation using the femur of Austrians born in the 19th to the middle of the 20th century. *Anthropologischer Anzeiger*, *72*(1), 117–127.

Kemkes, A., & Göbel, T. (2006). Metric assessment of the mastoid triangle for sex determination: A validation study. *Journal of Forensic Sciences*, *51*(5), 985–989.

Kemkes-Grottenhaler, A. (2005). Sex determination by discriminant analysis: An evaluation of the reliability of patella measurements. *Forensic Science International*, *147*(2–3), 129–133.

Kenyhercz, M. W., Klales, A. R., Stull, K. E., McCormick, K. A., & Cole, S. J. (2017). Worldwide population variation in pelvic sexual dimorphism: A validation and recalibration of the Klales et al. method. *Forensic Science International*, *277*. 259.e1–259.e8.

Kim, D. I., Lee, U. Y., Park, D. K., Kim, Y. I., Han, K. H., Kim, K. H., et al. (2006). Morphometrics of the hyoid bone for human sex determination from digital photographs. *Journal of Forensic Sciences*, *51*(5), 979–984.

Klales, A. R., Ousley, S. D., & Vollner, J. M. (2012). A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *American Journal of Physical Anthropology*, *149*(1), 104–114.

Kotěrová, A., Velemínská, J., Dupej, J., Brzobohatá, H., Pilný, A., & Brůžek, J. (2017). Disregarding population specificity: its influence on the sex assessment methods from the tibia. *International Journal of Legal Medicine*, *131*(1), 251–261.

Kranioti, E. F., Bastir, M., Sánchez-Meseguer, A., & Rosas, A. (2009). A geometric-morphometric study of the Cretan humerus for sex identification. *Forensic Science International*, *189*(1–3). 111-e1.

Kranioti, E. F., Iscan, M. Y., & Michalodimitrakis, M. (2008). Craniometric analysis of the modern Cretan population. *Forensic Science International*, *180*(2–3). 110.e1-110.e5.

Kranioti, E. F., & Michalodimitrakis, M. (2009). Sexual dimorphism of the humerus in contemporary cretans—A population-specific study and a review of the literature. *Journal of Forensic Sciences*, *54*(5), 996–1000.

Kranioti, E. F., Nathena, D., & Michalodimitrakis, M. (2011). Sex estimation of the Cretan humerus: A digital radiometric study. *International Journal of Legal Medicine*, *125*(5), 659–667.

Lovell, N. C. (1989). Test of Phenice's technique for determining sex from the *os pubis*. *American Journal of Physical Anthropology*, *79*(1), 117–120.

Mahfouz, M., Badawi, A., Merkl, B., Abdel Fatah, E. E., Pritchard, E., Kesler, K., et al. (2007). Patella sex determination by 3D statistical shape models and nonlinear classifiers. *Forensic Science International*, *173* (2–3), 161–170.

Meeusen, R. A., Christensen, A. M., & Hefner, J. T. (2015). The use of femoral neck axis length to estimate sex and ancestry. *Journal of Forensic Sciences*, *60*(5), 1300–1304.

Michiue, T., Hishmat, A. M., Oritani, S., Miyamoto, K., Amin, M. F., Ishikawa, T., et al. (2018). Virtual computed tomography morphometry of the patella for estimation of sex using postmortem Japanese adult data in forensic identification. *Forensic Science International*, *285*, 206.e1–206.e6.

Murail, P., Bruzek, J., Houët, F., & Cunha, E. (2005). DSP: A tool for probabilistic sex diagnosis using worldwide variability in hip-bone measurements. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, *17*(3–4), 167–176.

Osunwoke, E. A., Gwunireama, I. U., Orish, C. N., Ordu, K. S., & Ebowe, I. (2010). A study of sexual dimorphism of the human sternum in the southern Nigerian population. *Journal of Applied Biosciences*, *26*, 1636–1639.

Papaioannou, V. A., Kranioti, E. F., Joveneaux, P., Nathena, D., & Michalodimitrakis, M. (2012). Sexual dimorphism of the scapula and the clavicle in a contemporary Greek population: Applications in forensic identification. *Forensic Science International*, *217*(1–3). 231.e1–231.e7.

Papaloucas, C., Fiska, A., & Demetriou, T. (2008). Sexual dimorphism of the hip joint in Greeks. *Forensic Science International*, *179*(1). 83.e1–83.e3.

Phenice, T. W. (1969). A newly developed visual method for sexing the *os pubis*. *American Journal of Physical Anthropology*, *30*(2), 297–301.

Purkait, R., & Chandra, H. (2002). Sexual dimorphism in femora: An Indian Study. *Forensic Science Communications*, *4*(3).

Robinson, M. S., & Bidmos, M. A. (2008). An assessment of the accuracy of discriminant function equations for sex determination of the femur and tibia from a South African population. *Forensic Science International*, *206*(1–3), 67–72.

Rogers, T., & Saunders, S. (1994). Accuracy of sex determination using morphological traits of the human pelvis. *Journal of Forensic Sciences*, *39*(4), 1047–1056.

Rowbothan, S. K. (2016). Anthropological estimation of sex. In S. Blau & D. H. Ubelaker (Eds.), *Handbook of forensic anthropology and archaeology* (2nd ed., pp. 261–272). New York: Routledge.

Sharma, M., Gorea, R. K., Gorea, A., & Abuderman, A. (2016). A morphometric study of the human mandible in the Indian population for sex determination. *Egyptian Journal of Forensic Sciences*, *6*(2), 165–169.

Silventoinen, K. (2003). Determinants of variation in adult body height. *Journal of Biosocial Science*, *35*(2), 263–285.

Singh, J., Pathak, R. K., & Sing, D. (2012). Morphometric sex determination from various sternal widths of Northwestern Indian sternums collected from autopsy cadavers: A comparison of sexing methods. *Egyptian Journal of Forensic Sciences*, *2*(1), 18–28.

Soni, G., Dhall, U., & Chhabra, S. (2010). Determination of sex from femur: Discriminant analysis. *Journal of the Anatomical Society of India*, *59*(2), 216–221.

Srivastava, R., Saini, V., Rai, R. K., Pandey, S., & Tripahi, S. K. (2012). A study of sexual dimorphism in the femur among North Indians. *Journal of Forensic Sciences*, *57*(1), 19–23.

Stewart, T. D. (1979). *Essentials of forensic anthropology: Especially as developed in the United States*. IL: Bannerstone House.

Steyn, M., & İşcan, M. Y. (1997). Sex determination from the femur and tibia in South African whites. *Forensic Science International*, *90*(1–2), 111–119.

Steyn, M., & İşcan, M. Y. (1999). Osteometric variation in the humerus: Sexual dimorphism in South Africans. *Forensic Science International*, *106*(2), 77–85.

Steyn, M., & İşcan, M. Y. (2008). Metric sex determination from the pelvis in modern Greeks. *Forensic Science International*, *179*(1), 86–e1.

Torimitsu, S., Makino, Y., Saitoh, H., Sakuma, A., Ishii, N., Yajima, D., et al. (2015). Morphometric analysis of sex differences in contemporary Japanese pelves using multidetector computed tomography. *Forensic Science International*, *257*. 530.e1–530.e7.

Torimitsu, S., Makino, Y., Saitoh, H., Sakuma, A., Ishii, N., Yajima, D., et al. (2016a). Sex estimation based on scapula analysis in a Japanese population using multidetector computed tomography. *Forensic Science International*, *262*. 285.e1–285.e5.

Torimitsu, S., Makino, Y., Saitoh, H., Sakuma, A., Ishii, N., Yajima, D., et al. (2016b). Sexual determination based on multidetector computed tomographic measurements of the second cervical vertebra in a contemporary Japanese population. *Forensic Science International*, *266*. 588.e1–588.e6.

Torimitsu, S., Makino, Y., Saitoh, H., Sakuma, A., Ishii, N., Yajima, D., et al. (2018). Sex assessment based on clavicular measurements in a modern Japanese population using multidetector computed tomography. *Forensic Science International*, *285*, 207.e1–207.e5.

Ubelaker, D. H. (1992). Hyoid fracture and strangulation. *Journal of Forensic Sciences*, *37*(5), 1216–1222.

Ubelaker, D. H. (1999). *Human skeletal remains: excavation, analysis, interpretation* (3rd ed.). Washington, DC: Taraxacum.

Ubelaker, D. H. (2009). Historical development of forensic anthropology: Perspective from the United States. In S. Blau & D. H. Ubelaker (Eds.), *Handbook of forensic anthropology and archaeology* (pp. 76–86). Walnut Creek, CA: Left Coast Press.

Ubelaker, D. H. (2014). Osteology reference collections. In C. Smith (Ed.), *Encyclopedia of global archaeology* (pp. 5632–5641). New York, NY: Springer Science + Business Media.

Ubelaker, D. H., & DeGaglia, C. M. (2017). Population variation in skeletal sexual dimorphism. *Forensic Science International*, *278*. 407.e1–407.e7.

Ubelaker, D. H., & Volk, C. G. (2002). A test of the Phenice method for the estimation of sex. *Journal of Forensic Sciences*, *47*(1), 19–24.

CHAPTER 18

Secular change

Natalie R. Langley[a], **Richard L. Jantz**[b]
[a]Department of Laboratory Medicine and Pathology, Division of Anatomic Pathology, Mayo Clinic, Scottsdale, AZ, United States
[b]Department of Anthropology, University of Tennessee-Knoxville, Knoxville, TN, United States

Introduction

Secular changes are biological changes that occur over decades or generations, purportedly due to environmental factors (Roche, 1979). Two well-documented secular trends in global populations are increasing stature and earlier age of menarche (Fredriks, Van Burren, & Burgmeijer, 2000; Kim et al., 2008). Secular changes are more likely precipitated by eliminating growth-inhibiting factors (e.g., nutritional stress, environmental stresses, and disease) rather than introducing growth-stimulating factors (Malina, 1979). Contributing variables are associated with improvements in living conditions during the late 19th and early 20th centuries. Overall health and life expectancy were enhanced by improved sanitation, nutrition, and public health; elimination of epidemic diseases; reduced infant mortality; and technological advances.

While the precipitating factors of secular change are not debated (i.e., environmental factors), the mechanism is not fully understood. As a result, plasticity continues to be cited as a primary mechanism, suggesting phenotypic change without genetic change (Langley, Jantz, & Ousley, 2016). However, epigenetic research demonstrates that environmental changes can alter gene expression without altering the DNA sequence (e.g., through DNA methylation or histone modification). The resulting phenotypic response can be passed on to future generations (Dias & Ressler, 2014; Guth et al., 2013). Dramatic changes in the past 200 years suggest that epigenetic changes have modified the human skeletal structure, although they cannot be specified presently (Langley et al., 2016). Another genetic factor that may have precipitated changes in skeletal form is reduced genetic isolation and the concomitant increase in heterozygosity (Boldsen, 1995; Hulse, 1964). Furthermore, selection cannot be eliminated as a mechanism solely on the grounds that not enough time has elapsed for significant genetic change to occur since rapid evolution has been documented in many organisms, including birds (Brown & Bomberger Brown, 2013), rodents (Harris, Munshi-South, Obergfell, & O'Neill, 2013; Pergams & Lawler, 2009), lizards (Herrel et al., 2008), dogs (Drake & Klingenberg, 2008), and humans (Milot et al., 2011). Stulp, Barrett, Tropf, and Mills (2015) have specifically found support for the hypothesis that the Dutch, now the tallest population in the world, is partly due to selection.

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00018-3

Environmental influences cannot erase deep-seated genetic differences between populations, but the undeniable evidence of secular change emphasizes the importance of evaluating the short-term effects of environment on skeletal morphology (Kouchi, 2004), and provides a cautionary tale for the selection of methods and reference samples in forensic contexts. Secular change in bony dimensions and skeletal maturation prescribe that the medical and forensic communities use data from modern populations to deduce information about growth, health, and biological traits (Langley & Cridlin, 2016).

Evidence of secular change

Accelerated maturation, evidenced by a steady decline in menarcheal age, has been documented extensively in populations around the globe. The link between environmental conditions and age at menarche has been demonstrated in numerous studies. Malina (1979) reported a delay in maturation during the Industrial Revolution due to disease, nutritional, and social stresses associated with overcrowded cities. Shortly thereafter, improvements in environmental and nutritional quality led to accelerated growth. Since the 1920s, Europeans have experienced rather stable caloric intakes and reduced caloric expenditures and, consequently, increases in body weights. The only reported negative secular trend in menarcheal age was during World War II, but this was minor and temporary, and acceleration resumed once social and economic conditions were restored to pre-war levels.

Age at menarche in western industrialized populations has decreased by 4–6 months per decade over the last five decades (Fredriks et al., 2000; Malina, 1979), and pubertal onset in American females occurs as early as 8–10 years of age (Fredriks et al., 2000; Herman-Giddens et al., 1997; Malina, 1979; Morrison et al., 1994). Decreases in menarcheal age have been documented in Japan (Hoshi & Kouchi, 1981), South Korea (Hwang, Shin, Frongillo, Shin, & Jo, 2003), China (Huen et al., 1997; Leung, Lau, Xu, & Tse, 1996; Lin, Chen, Su, Xiao, & Ye, 1992; Low, Kung, & Leong, 1982; Low, Kung, Leong, & Hsu, 1981; So & Yen, 1992), India (Bagga & Kulkarni, 2000), Venezuela (Farid-Coupal, Contreras, & Castellano, 1981), Poland (Laska-Mierzejewska, Milicer, & Piechaczek, 1982), Finland (Rimpela & Rimpela, 1993), Belgium (Vercauteren & Susanne, 1985), the Netherlands (Fredriks et al., 2000; Wellens, Malina, Beunen, & Lefevre, 1990), Britain (Cameron, 1979), Spain (Prado, 1984), Sweden (Liu, Wikland, & Karlberg, 2000), France (La Rocherbrochard, 2000), and Austria (Kralj-Cercek, 1956). Although the precise age of pubertal onset is more difficult to detect in males, a marked decrease in the age of voice-breaking in males has been noted (Daw, 1970; La Rocherbrochard, 2000; Taranger, Engström, Lichtenstein, & Svennberg-Redegren, 1976).

Since sexual maturation is closely related to skeletal maturation, acceleration in pubertal onset means acceleration in skeletal maturation (Maresh, 1972). Data from the Polish population showed acceleration by 0.22–0.66 years per decade in epiphyseal fusion of the hand

and wrist (Himes, 1984). A similar trend was detected in southern Chinese girls in Hong Kong using radiographic data from the hand and wrist (So & Yen, 1990). Crowder and Austin (2005) reported that contemporary North American adolescents show advanced union of the tibia and fibula compared to earlier studies. In addition, the medial clavicle epiphysis was found to commence fusion 3–4 years earlier in late-20th-century Americans compared to the early and mid-20th century (Langley-Shirley & Jantz, 2010). This is also the case for the spheno-occipital synchondrosis (Shirley & Jantz, 2011), which closes about 2 years earlier than in the 19th century, an event that mirrors earlier puberty and peak height velocity in stature.

Secular changes in skeletal maturation affect adult morphology, including overall stature, long bone length and proportions, and cranial size and shape. Secular change in stature has been documented extensively (e.g., Bogin, Smith, Orden, Varela Silva, & Loucky, 2002; Eveleth & Tanner, 1990; Floud, 1994; Floud, Harris, & Hong, 2011; Fogel, Engerman, & Floud, 1983; Mokyr & Grada, 1994; Steckel, 1994; Trotter & Gleser, 1951). Fogel (1984) reported a slow gain in stature prior to 1850, a decrease from 1850 to 1900, and then a post-1900 recovery. Importantly, Meadows and Jantz (1995) note that stature is lowest during the decades corresponding to birth dates of individuals in the Terry collection, the skeletal collection upon which Trotter and Gleser (1952, 1958) derived some of their stature estimation formulae.

Overall, long bones have become more linear, narrow, and gracile. Distal elements have increased in length more than proximal elements, resulting in increased brachial and crural indices, and secular change in the lower limb is more pronounced than the upper limb (Jantz, Meadows Jantz, & Devlin, 2016; Meadows & Jantz, 1995; Meadows Jantz & Jantz, 1999). The shape of the femoral midshaft has become more oval in cross-section compared to a round cross-sectional shape in the mid-19th century. This change is due primarily to medio-lateral narrowing rather than antero-posterior elongation, likely on account of reduced physical activity among modern Americans (Jantz et al., 2016; Wescott & Zephro, 2016).

The question of whether secular changes in long bone length and proportions can be explained as an allometric response to increasing stature was examined by Meadows and Jantz (1995). Distal bones were positively allometric with stature, resulting in relatively longer distal bones in taller individuals. A re-examination of this question revealed the somewhat surprising result that the humerus is becoming relatively shorter in recent Euro-Americans and is driving the increase in the brachial index (Jantz & Meadows Jantz, 2017). They conclude that allometry alone cannot explain changing limb proportions, and that the unique American environment is in the process of restructuring allometric relationships.

Langley and Cridlin (2016) also documented a decrease in the maximum length of the clavicle after the mid-1900s, which is consistent with the pattern of overall narrowing in bones of the lower extremity. Cridlin (2016) reported a decrease in the robusticity of the femoral head in American whites, indicated by a significant decrease in maximum femoral head diameter after 1920. Changes in pelvic morphology are also consistent with the trend toward a more narrow and linear skeletal form in modern Americans (e.g., decreased pelvic breadth and increased innominate height) (Driscoll, 2012; Klales, 2016). Klales (2016) observed a more gracile pelvic form in modern females and males compared to historic (Hamann-Todd, HT) females (e.g., a narrower ischio-pubic ramus and pronounced subpubic concavity in females and a less convex subpubic angle in modern males compared to HT males).

Cranial changes over the past 200 years generally follow the overall narrowing and lengthening patterns observed in the postcranial skeleton, including narrower vault and face, higher vault, and longer cranial base (Jantz, 2001; Jantz & Meadows Jantz, 2000, 2016; Jantz & Wescott, 2002; Jonke et al., 2007; Weisensee & Jantz, 2016; Wescott & Jantz, 2005). Cranial size has increased slightly overall due to the fact that the changes in vault height and length are more pronounced than the decrease in breadth. Jantz and Meadows Jantz (2016) showed a strong correlation between changes in cranial shape and stature, suggesting that the cranium is responding to similar forces (e.g., improvements in health and nutrition, increasing wealth, and decreases in mortality and morbidity).

Implications for estimating sex from skeletal data

While there is no debating that the size and shape of skeletons have changed drastically over the past two centuries, the implications of this change for estimating sex from skeletal features and dimensions have not been investigated thoroughly. The importance of appropriate reference samples has been elucidated in regard to skeletal age-at-death estimation of adults and subadults (Boldsen, Milner, Konigsberg, & Wood, 2002; Kimmerle, Konigsberg, Jantz, & Baraybar, 2008; Konigsberg, Herrmann, Wescott, & Kimmerle, 2008; Langley-Shirley & Jantz, 2010; Schmeling, Schulz, Danner, & Rösing, 2006). We also know that significant population differences in skeletal dimensions and body proportions affect the accuracy of sex estimates from long bone measurements (Calcagno, 1981; Spradley, Jantz, Robinson, & Peccerelli, 2008; Thieme & Schull, 1957). The question is whether the magnitude of change over the last 200 years has been significant enough to affect the accuracy of sex estimates to the same extent as between-population differences in skeletal morphology.

Morphological sex estimation

Walker (2008) presented a method of sexing skulls using five visually assessed traits. It could achieve correct sex estimates approaching 90% on the samples he used. His European-derived samples consist of the English St. Brides collection, individuals with birth years in the last half of the 18th century, and the Terry Anatomical and HT

Collections, containing individuals born in the last half of the 19th century. Thus, the birth cohorts of the two samples are separated by 100 years. Neither of these samples would be applicable to modern Euro-Americans if there has been significant secular change in the traits Walker used. Two of Walker's most effective traits, glabella size and mastoid size, have been shown to exhibit secular increase in size since the 19th century using measurements glabella subtense and mastoid height (Manthey, Jantz, Bohnert, & Jellinghaus, 2016).

We further evaluate secular change in Walker's traits by comparing his 18th- and 19th-century trait frequencies to each other and to a preliminary sample ($n = 102$) from the University of Tennessee (UT)-donated collection scored by Meadows Jantz and Langley (unpublished data). Table 1 shows the results of a likelihood Chi-square test comparing glabella and mastoid scores for the samples representing three centuries. Eighteenth-century skulls are significantly different from those of the 19th century for both traits in males and glabella in females. For 19th vs. 20th centuries, males differ for mastoid scores, and females for glabella scores.

Changes over the three centuries can be shown by plotting the mean scores on century. Fig. 1 shows that both glabella and mastoid scores have increased, in males more than in females, resulting in an apparent increase in sex dimorphism. Consequently, using Walker's equations to sex modern Euro-American crania will cause too many females to classify as males. We tested Walker's equation using glabella and mastoid scores from the UT-donated sample mentioned above. Using his own samples, Walker's equation for his pooled English/American sample classifies 85.4% and 82.9% of males and females correctly. However, Table 2 shows that all 20th-century males are correctly classified, but only 65.6% of females. Horbaly (2017) found a similar pattern on modern Americans.

The results presented above have clearly demonstrated that Walker's (2008) equations cannot be used on modern Euro-Americans. Their applicability to African Americans remains to be tested. Walker's approach may well be suitable in forensic applications, but not without recalibrating using modern data.

Table 1 Comparison of frequencies of Walker scores for Walker's 18th-century English, 19th-century Euro-Americans (Walker, 2008), and 20th-century Euro-Americans (Meadows Jantz and Langley, unpublished). N = sample size.

Comparison		**Glabella**		**Mastoid**	
Males	**N_1 vs. N_2**	**Likelihood χ^2**	**Probability**	**Likelihood χ^2**	**Probability**
18th vs. 19th	43/61	70.26	<0.0001	51.14	<0.0001
19th vs. 20th	61/38	2.74	0.60	17.58	0.002
Females					
18th vs. 19th	35/52	15.45	0.002	4.86	0.183
19th vs. 20th	52/64	17.93	0.001	3.64	0.303

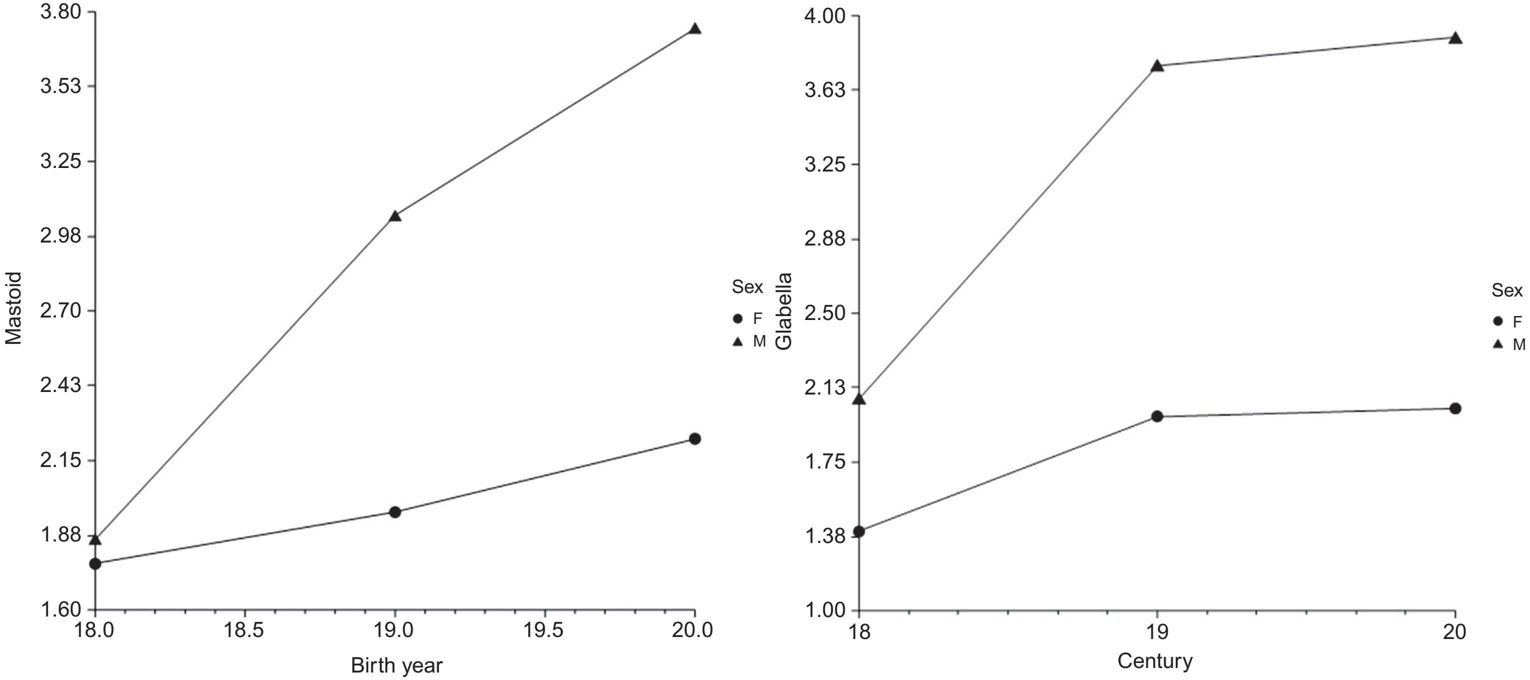


Fig. 1 Mean Walker (2008) glabella and mastoid scores for 18th-, 19th-, and 20th-century samples, showing secular change in these two traits.

Table 2 Classification of modern individuals from the UT-donated collection using Walker's (2008) discriminant function for glabella and mastoid.

		Into sex		
From sex	***N***	**Male**	**Female**	**% correct**
Male	38	38	0	100
Female	64	22	42	65.63

Klales' (2016) analysis of non-metric pelvic traits showed that scores of three pelvic traits for modern female skeletons were consistently lower than those of historic (early-20th-century) samples, indicating a secular trend toward a more gracile pelvic form in modern females. Trait expression in males was more consistent through time, with the robust expression being the most common. The three traits were revised descriptions (Klales, Ousley, & Vollner, 2012) of Phenice's (1969) traits: subpubic contour, medial aspect of ischiopubic ramus, and ventral arc. Klales et al. (2012) developed the revised traits on HT collection pelves and tested the revised descriptions on UT-donated collection skeletons. Ordinal logistic regression bore out the presence of morphological variation between the temporal samples. The traits were able to assign 84.1% of HT sample pelves to the correct temporal population, but only 53.3% of the modern sample; classification was lower for males than females. However, the method performed acceptably at sex estimation on an independent test sample of modern American skeletons (86.2% correctly sexed), suggesting that the magnitude of change in pelvic form has not appreciably affected the accuracy of non-metric pelvic sexing using the Klales et al. (2012) method.

Metric sex estimation

Using measurements to estimate sex with discriminant function analysis was introduced as early as 1957 (Thieme, 1957; Thieme & Schull, 1957). The most influential of these early attempts to estimate sex using measurements is that of Giles and Elliot (1963), which was based on the Terry and Todd collections. Giles and Elliot's discriminant functions have been used extensively by forensic anthropologists until recently. It is, therefore, worth considering how secular change may have influenced Giles and Elliot's sexing criteria. Fig. 2 shows the distribution of Giles and Elliot's scores in relation to birth years for 19th- and 20th-century crania, using their function #4, for whites. It shows that the scores exhibit progressive increase through time. The dashed line represents the Giles and Elliot's sectioning point. It divides the sexes approximately evenly during the 19th century, but by mid- to late 20th century, the section point has approached the female mean such that almost all males are correctly classified and too many females are classified as males. Presumably, changes in long bone lengths and pelvic dimensions would shift sectioning points for discriminant functions derived from postcranial

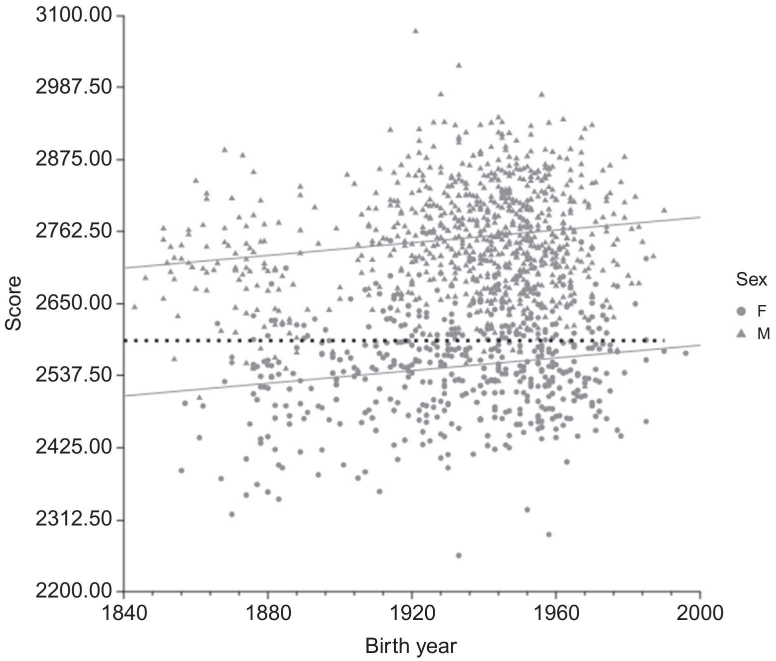


Fig. 2 Plot of secular changes in Giles and Elliot's (1963) sex discriminant function (#4) score. It shows that the average score increases with year of birth. The dashed line is Giles and Elliot' sectioning point. It shows that in the 20th century, especially after about 1940, the function will classify too many females as males.

measurements as well. While the overall accuracy rates of these functions may remain relatively high because postcranial dimensions are powerful sex estimators, individuals near the sectioning point will be inevitably misclassified.

Conclusion

Franz Boas (1911) investigated the changes in head shape of immigrants and their children over a century ago. His anthropometric study called attention to the morphological plasticity of human biology in the context of novel environments. While his conclusions have undergone some modification (Jantz & Logan, 2010), his observations introduced a concept that has important implications for forensic anthropology practice today, particularly in light of the reference samples from which many methods have been developed. Multiple factors are responsible for secular changes in skeletal form, and biologists have no reason to assume that human populations will reach stasis.

The examples provided in this chapter illustrate that appropriate reference samples provide the most accurate results for sex estimation from metric and morphological skeletal features. Skeletal form has changed over the past century; therefore, methods derived from historical reference samples should be avoided in modern forensic contexts unless proved as acceptable by validation studies. If the goal is to facilitate the identification of unknown remains by establishing the most accurate biological profile from the available skeletal data, then best practice necessitates factoring in the effects of secular change on discriminant function sectioning points.

References

Bagga, A., & Kulkarni, S. (2000). Age at menarche and secular trend in Maharashtrian (Indian) girls. *Acta Biologica Szegediensis*, *44*(1-4), 53–57.

Boas, F. (1911). *Changes in bodily form of descendants of immigrants. Final Report.* Washington, D.C.: Government Printing Office.

Bogin, B., Smith, P., Orden, A. B., Varela Silva, M. I., & Loucky, J. (2002). Rapid change in height and body proportions of Maya American children. *American Journal of Human Biology*, *14*(6), 753–761. https://doi.org/10.1002/ajhb.10092.

Boldsen, J. L. (1995). The place of plasticity in the secular trend for male stature: An analysis of Danish biological population history. In C. G. N. Mascie-Taylor & B. Bogin (Eds.), *Human variability and plasticity* (pp. 75–90). Cambridge: Cambridge University Press.

Boldsen, J. L., Milner, G. R., Konigsberg, L. W., & Wood, J. W. (2002). Transition analysis: A new method for estimating age from skeletons. In R. D. Hoppa & J. W. Vaupel (Eds.), *Paleodemography: Age distributions from skeletal samples* (pp. 73–106). Cambridge: Cambridge University Press.

Brown, C. R., & Bomberger Brown, M. (2013). Where has all the road kill gone? *Current Biology*, *23*(6), R233–R234. https://doi.org/10.1016/j.cub.2013.02.023.

Calcagno, J. M. (1981). On the applicability of sexing human skeletal material by discriminant function analysis. *Journal of Human Evolution*, *10*(2), 189–198.

Cameron, N. (1979). The growth of London England UK school children 1904–1966: An analysis of secular trend and intra-county variation. *Annals of Human Biology*, *6*(6), 505–526.

Cridlin, S. (2016). Piecewise regression analysis of secular change in the maximum femoral vertical head diameter of American white males and females. *Human Biology*, *88*(1), 47–55.

Crowder, C., & Austin, D. (2005). Age ranges of epiphyseal fusion in the distal tibia and fibula of contemporary males and females. *Journal of Forensic Sciences*, *50*(5), 1001–1007.

Daw, S. (1970). Age of boys' puberty in Leipzig, 1727-49, as indicated by voice breaking in J.S. Bach's choir members. *Human Biology*, *42*(1), 87–89.

Dias, B. G., & Ressler, K. J. (2014). Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nature Neuroscience*, *17*(1), 89–96. https://doi.org/10.1038/nn.3594.

Drake, A. G., & Klingenberg, C. P. (2008). The pace of morphological change: Historical transformation of skull shape in St Bernard dogs. *Proceedings of the Biological Sciences*, *275*(1630), 71–76. https://doi.org/10.1098/rspb.2007.1169.

Driscoll, K. R. (2012). Secular change in the modern human pelvis. In *Paper presented at the American Association of Physical Anthropologists annual meeting, Portland, OR.*

Eveleth, P., & Tanner, J. (1990). *Worldwide variation in human growth* (2nd ed.). Cambridge: Cambridge University Press.

Farid-Coupal, N., Contreras, M. L., & Castellano, H. M. (1981). The age at menarche in Carabobo, Venezuela with a note on the secular trend. *Annals of Human Biology*, *8*(3), 283–288.

Floud, R. (1994). The heights of Europeans since 1975: A new source for European economic history. In J. Komlos (Ed.), *Stature, living standards and economic development* (pp. 39–59). Chicago, IL: University of Chicago Press.

Floud, R. R., Harris, B., & Hong, S. C. (2011). *The changing body: Health, nutrition, and human development in the Western World since 1700.* Cambridge, UK: Cambridge University Press.

Fogel, R. W. (1984). Nutrition and the decline in mortality since 1700: Some preliminary findings. In S. L. Engerman & R. E. Gallman (Eds.), *Long-term factors in American economic growth* (p. 439). Chicago, IL: University of Chicago Press.

Fogel, R. W., Engerman, S. L., & Floud, R. (1983). Secular changes in American and British stature and nutrition. *Journal of Interdisciplinary History*, *14*(2), 445–481.

Fredriks, A., Van Burren, S., & Burgmeijer, R. (2000). Continuing positive secular growth change in the Netherlands 1955–1997. *Pediatric Research*, *47*(3), 316–323.

Giles, E., & Elliot, O. (1963). Sex determination by discriminant function analysis of crania. *American Journal of Physical Anthropology*, *21*, 53–68.

Guth, L. M., Ludlow, A. T., Witkowski, S., Marshall, M. R., Lima, L. C., Venezia, A. C., et al. (2013). Sex-specific effects of exercise ancestry on metabolic, morphological and gene expression phenotypes in multiple generations of mouse offspring. *Experimental Physiology*, *98*(10), 1469–1484. https://doi.org/10.1113/expphysiol.2012.070300.

Harris, S. E., Munshi-South, J., Obergfell, C., & O'Neill, R. (2013). Signatures of rapid evolution in urban and rural transcriptomes of white-footed mice (*Peromyscus leucopus*) in the New York metropolitan area. *PLoS One*, *8*(8), e74938. https://doi.org/10.1371/journal.pone.0074938.

Herman-Giddens, M. E., Slora, E. J., Wasserman, R. C., Bourdony, C. J., Bhapkar, M. V., Koch, G. G., et al. (1997). Secondary sexual characteristics and menses in young girls seen in office practice: A study from the Pediatric Research in Office Settings Network. *Pediatrics*, *99*(4), 505–512.

Herrel, A., Huyghe, K., Vanhooydonck, B., Backeljau, T., Breugelmans, K., Grbac, I., et al. (2008). Rapid large-scale evolutionary divergence in morphology and performance associated with exploitation of a different dietary resource. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(12), 4792–4795. https://doi.org/10.1073/pnas.0711998105.

Himes, J. (1984). Early hand-wrist atlas and its implications for secular change in bone age. *Annals of Human Biology*, *11*(1), 71–75.

Horbaly, H. E. (2017). The influence of stature on sex estimation. In *Paper presented at the American Academy of Forensic Sciences annual meeting, New Orleans, LA*.

Hoshi, H., & Kouchi, M. (1981). Secular trend of the age of menarche of Japanese girls with special regard to the secular acceleration of the age at peak height velocity. *Human Biology*, *53*(4), 593–598.

Huen, K. F., Leung, S. S. F., Lau, J. T. F., Cheung, A. Y. K., Leung, N. K., & Chiu, M. C. (1997). Secular trend in the sexual maturation of Southern Chinese girls. *Acta Paediatrica*, *86*(10), 1121–1124.

Hulse, F. S. (1964). Exogamy and heterosis. *Yearbook of Physical Anthropology*, *9*, 240–257.

Hwang, J., Shin, C., Frongillo, E., Shin, K., & Jo, I. (2003). Secular trend in age at menarche for South Korean women born between 1920 and 1986: the Ansan Study. *Annals of Human Biology*, *30*(4), 434–442.

Jantz, R. L. (2001). Cranial change in Americans: 1850–1975. *Journal of Forensic Sciences*, *46*(4), 784–787.

Jantz, R. L., & Logan, M. H. (2010). Why does head form change in children of immigrants? A reappraisal. *American Journal of Human Biology*, *22*(5), 702–707. https://doi.org/10.1002/ajhb.21070.

Jantz, R. L., & Meadows Jantz, L. (2000). Secular change in craniofacial morphology. *American Journal of Human Biology*, *12*(3), 327–338. https://doi.org/10.1002/(sici)1520-6300(200005/06)12:3<327::aid-ajhb3>3.0.co;2-1.

Jantz, R. L., & Meadows Jantz, L. (2016). The remarkable change in Euro-American cranial shape and size. *Human Biology*, *88*(1), 56–64.

Jantz, R. L., & Meadows Jantz, L. (2017). Limb bone allometry in modern Euro-Americans. *American Journal of Physical Anthropology*, *163*(2), 252–263. https://doi.org/10.1002/ajpa.23203.

Jantz, R. L., Meadows Jantz, L., & Devlin, J. (2016). Secular changes in the postcranial skeleton of American whites. *Human Biology*, *88*(1), 65–75.

Jantz, R. L., & Wescott, D. J. (2002). Assessing craniofacial secular change in American whites and blacks using geometric morphometry. *American Journal of Physical Anthropology Supplement*, *34*, 90.

Jonke, E., Prossinger, H., Bookstein, F. L., Schaefer, K., Bernhard, M., & Freudenthaler, J. W. (2007). Secular trends in the facial skull from the 19th century to the present, analyzed with geometric morphometrics. *American Journal of Orthodontics and Dentofacial Orthopedics*, *132*(1), 63–70. https://doi.org/10.1016/j.ajodo.2005.08.040.

Kim, J. Y., Oh, I. H., Lee, E. Y., Choi, K. S., Choe, B. K., Yoon, T. Y., et al. (2008). Anthropometric changes in children and adolescents from 1965 to 2005 in Korea. *American Journal of Physical Anthropology*, *136*(2), 230–236.

Kimmerle, E. H., Konigsberg, L. W., Jantz, R. L., & Baraybar, J. P. (2008). Analysis of age-at-death estimation through the use of pubic symphyseal data. *Journal of Forensic Sciences*, *53*(3), 558–568. https://doi.org/10.1111/j.1556-4029.2008.00711.x.

Klales, A. R. (2016). Secular change in morphological pelvic traits used for sex estimation. *Journal of Forensic Sciences*, *61*(2), 295–301. https://doi.org/10.1111/1556-4029.13008.

Klales, A. R., Ousley, S. D., & Vollner, J. M. (2012). A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *American Journal of Physical Anthropology*, *149*(1), 104–114. https://doi.org/10.1002/ajpa.22102.

Konigsberg, L. W., Herrmann, N. P., Wescott, D. J., & Kimmerle, E. H. (2008). Estimation and evidence in forensic anthropology: age-at-death. *Journal of Forensic Sciences*, *53*(3), 541–557. https://doi.org/10.1111/j.1556-4029.2008.00710.x.

Kouchi, M. (2004). Secular changes in the Japanese head form viewed from somatometric data. *Anthropological Science*, *112*, 41–52.

Kralj-Cercek, L. (1956). The influence of food, body build, and social origin on the age at menarche. *Human Biology*, *28*(4), 393–406.

La Rocherbrochard, E. (2000). Age at puberty of girls and boys in France: Measurements from a survey on adolescent sexuality. *Population: An English Selection*, *12*, 51–79.

Langley, N. R., & Cridlin, S. (2016). Changes in clavicle length and maturation in Americans: 1840–1980. *Human Biology*, *88*(1), 76–83.

Langley, N. R., Jantz, R. L., & Ousley, S. D. (2016). The effect of novel environments on modern American skeletons. *Human Biology*, *88*(1), 5–13.

Langley-Shirley, N. R., & Jantz, R. L. (2010). A Bayesian approach to age estimation in modern Americans from the clavicle. *Journal of Forensic Sciences*, *55*(3), 571–583. https://doi.org/10.1111/j.1556-4029.2010.01089.x.

Laska-Mierzejewska, T., Milicer, H., & Piechaczek, H. (1982). Age at menarche and its secular trend in urban and rural girls in Poland. *Annals of Human Biology*, *9*(3), 227–234.

Leung, S., Lau, J., Xu, Y., & Tse, L. (1996). Secular changes in standing height, sitting height and sexual maturation of Chinese—The Hong Kong growth study, 1993. *Annals of Human Biology*, *23*(4), 297–306.

Lin, W., Chen, A., Su, J., Xiao, J., & Ye, G. (1992). Secular change in the growth and development of Han children in China. *Annals of Human Biology*, *19*(3), 249–265.

Liu, Y. X., Wikland, K. A., & Karlberg, J. (2000). New reference for the age at childhood onset of growth and secular trend in the timing of puberty in Swedish. *Acta Paediatrica*, *89*(6), 637–643.

Low, W., Kung, L., & Leong, J. (1982). Secular trend in the sexual maturation of Chinese girls. *Human Biology*, *54*(3), 539–552.

Low, W., Kung, L., Leong, J., & Hsu, L. (1981). The secular trend in the growth of southern Chinese girls in Hong Kong. *Zeitschrift für Morphologie und Anthropologie*, *72*(1), 77–88.

Malina, R. (1979). Secular changes in size and maturity: Causes and effects. *Monographs of the Society for Research in Child Development*, *44*(3-4), 59–102.

Manthey, L., Jantz, R. L., Bohnert, M., & Jellinghaus, K. (2016). Secular change of sexually dimorphic cranial variables in Euro-Americans and Germans. *International Journal of Legal Medicine*. https://doi.org/10.1007/s00414-016-1469-2.

Maresh, M. M. (1972). A 45 year investigation for secular changes in physical maturation. *American Journal of Physical Anthropology*, *36*(1), 103–110.

Meadows Jantz, L., & Jantz, R. L. (1999). Secular change in long bone length and proportion in the United States, 1800–1970. *American Journal of Physical Anthropology*, *110*, 57–67.

Meadows, L., & Jantz, R. L. (1995). Allometric secular change in the long bones from the 1800s to the present. *Journal of Forensic Sciences*, *40*(5), 762–767.

Milot, E., Mayer, F. M., Nussey, D. H., Boisvert, M., Pelletier, F., & Reale, D. (2011). Evidence for evolution in response to natural selection in a contemporary human population. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(41), 17040–17045. https://doi.org/10.1073/pnas.1104210108.

Mokyr, J. O., & Grada, C. (1994). The heights of the British and the Irish c. 1800–1815: Evidence from recruits to the East India Company's Army. In J. Komlos (Ed.), *Stature, living standards and economic development* (pp. 39–59). Chicago, IL: University of Chicago Press.

Morrison, J. A., Barton, B., Biro, F. M., Sprecher, D. L., Falkner, F., & Obarzanek, E. (1994). Sexual maturation and obesity in 9-year-old and 10-year-old black and white girls: The national heart and lung and blood institute growth and health study. *Journal of Pediatrics*, *124*(6), 889–895.

Pergams, O. R., & Lawler, J. J. (2009). Recent and widespread rapid morphological change in rodents. *PLoS One*, *4*(7), e6452. https://doi.org/10.1371/journal.pone.0006452.

Phenice, T. W. (1969). A newly developed visual method of sexing *os pubis*. *American Journal of Physical Anthropology*, *30*(2), 297.

Prado, C. (1984). Secular change in menarche in women in Madrid. *Annals of Human Biology*, *11*(2), 165–166.

Rimpela, A. H., & Rimpela, M. K. (1993). Towards an equal distribution of health? Socioeconomic and regional differences of the secular trend of the age of menarche in Finland from 1979 to 1989. *Acta Paediatrica*, *82*(1), 87–90.

Roche, A. (1979). Secular trends in human growth, maturation, and development. *Monographs of the Society for Research in Child Development*, *44*(3-4), 1–120.

Schmeling, A., Schulz, R., Danner, B., & Rösing, F. (2006). The impact of economic progress and modernization in medicine on the ossification of hand and wrist. *International Journal of Legal Medicine*, *120*, 121–126. https://doi.org/10.1007/s00414-005-0007-4.

Shirley, N. R., & Jantz, R. L. (2011). Spheno-occipital synchondrosis fusion in modern Americans. *Journal of Forensic Sciences*, *56*(3), 580–585. https://doi.org/10.1111/j.1556-4029.2011.01705.x.

So, L., & Yen, P. (1990). Secular trend in skeletal maturation in Southern Chinese girls in Hong Kong. *Zeitschrift für Morphologie und Anthropologie*, *78*(2), 145–154.

So, L., & Yen, P. (1992). Secular trend of menarcheal age in Southern Chinese girls. *Zeitschrift für Morphologie und Anthropologie*, *79*(1), 21–24.

Spradley, M. K., Jantz, R. L., Robinson, A., & Peccerelli, B. (2008). Demographic change and forensic identification: Problems in metric identification of hispanic skeletons. *Journal of Forensic Sciences*, *53*(1), 21–28.

Steckel, R. H. (1994). Heights and health in the United States, 1710–1950. In J. Komlos (Ed.), *Stature, living standards and economic development* (pp. 153–172). Chicago, IL: University of Chicago Press.

Stulp, G., Barrett, L., Tropf, F. C., & Mills, M. (2015). Does natural selection favour taller stature among the tallest people on earth? *Proceedings of the Biological Sciences*, *282*(1806), 20150211. https://doi.org/10.1098/rspb.2015.0211.

Taranger, J., Engström, I., Lichtenstein, H., & Svennberg-Redegren, I. (1976). Somatic pubertal development. *Acta Paediatrica Scandinavica Supplement*, *258*, 121–135.

Thieme, F. (1957). Sex in Negro skeletons. *Journal of Forensic Medicine*, *4*, 72–81.

Thieme, F., & Schull, W. (1957). Sex determination from the skeleton. *Human Biology*, *29*(3), 242–273.

Trotter, M., & Gleser, G. C. (1951). The effect of aging on stature. *American Journal of Physical Anthropology*, *9*, 311–324.

Trotter, M., & Gleser, G. C. (1952). Estimation of stature from long bones of American Whites and Negroes. *American Journal of Physical Anthropology*, *10*, 463–514.

Trotter, M., & Gleser, G. C. (1958). A re-evaluation of estimation of stature based on measurements of stature taken during life and of long bones after death. *American Journal of Physical Anthropology*, *16*, 79–123.

Vercauteren, M., & Susanne, C. (1985). The secular trend of height and menarche in Belgium: Are there any signs of a future stop. *European Journal of Pediatrics*, *144*(4), 306–309.

Walker, P. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, *136*(1), 39–50.

Weisensee, K. E., & Jantz, R. L. (2016). An examination of the differential effects of the modern epidemiological transition on cranial morphology in the United States and Portugal. *Human Biology*, *88*(1), 30–37.

Wellens, R., Malina, R., Beunen, G., & Lefevre, J. (1990). Age at menarche in Flemish girls: Current status and secular change in the 20th century. *Annals of Human Biology*, *17*(2), 145–152.

Wescott, D. J., & Jantz, R. L. (2005). Assessing craniofacial secular change in American blacks and whites using geometric morphometry. In D. E. Slice (Ed.), *Modern morphometrics in physical anthropology* (pp. 231–245). New York: Kluwer Academic/Plenum.

Wescott, D. J., & Zephro, L. (2016). Secular change in the femur diaphyseal biomechanical properties of American whites. *Human Biology*, *88*(1), 38–46.

CHAPTER 19

The effects of skeletal asymmetry on accurate sex classification

Stephanie J. Cole, Cortney N. Hulse, Kyra E. Stull
Department of Anthropology, University of Nevada-Reno, Reno, NV, United States

Introduction

Sex estimation is one of the most important aspects of the biological profile, as a reliable estimation of this parameter often allows for the application of more accurate sex-specific methods for the estimation of age, stature, and ancestry. Most sex estimation methods utilize bilateral traits or measurements of the skeleton. By convention, forensic anthropologists and bioarchaeologists typically develop methods using elements from the left side. However, in practice, the left side may not always be available for analysis (e.g., taphonomy, trauma, pathology) and if individuals are asymmetric in their bilateral traits or measurements, the right and left sides may yield contradictory sex estimates. Consequently, preferentially selecting one side over the other could introduce fundamental biases that compromise method accuracy. Despite this issue, research regarding the effects of asymmetry on accurate sex classification is lacking. Thus, the aims of this chapter are to provide a brief synopsis of what is currently known regarding the impact of asymmetry on sex classification and to introduce new research on the pelvis to gain a better understanding of the topic.

Defining asymmetry

There are three primary types of asymmetry in biological organisms: directional asymmetry, anti-symmetry, and fluctuating asymmetry (Van Valen, 1962). Directional asymmetry and anti-symmetry affect one side of the body more than the other, resulting in larger dimensions of the dominant side (Franks & Cabo, 2014; Palmer & Strobeck, 1986; Van Valen, 1962). Regarding directional asymmetry, the same side is consistently affected throughout the population. A classic example of directional asymmetry is the overwhelming occurrence of right-handed individuals compared with left-handed individuals in human populations (Graham & Özener, 2016). In cases of anti-symmetry, the observed asymmetry does not consistently favor one side, resulting in approximately half the population displaying right dominance and half displaying left dominance (Graham & Özener, 2016; Palmer, 1996). An example of anti-symmetry found in nature

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00019-5

is the signaling claw of male fiddler crabs. These animals use their larger claw to attract females regardless of whether the larger claw is expressed on the right or left side (Palmer & Strobeck, 1986). The result is an evenly distributed occurrence of dominant right and left expression throughout the species (Franks & Cabo, 2014; Graham & Özener, 2016). Both directional and anti-symmetry are presumed to be adaptive, resulting either from side preference in the case of directional asymmetry, or differential gene activity in the case of anti-symmetry (Franks & Cabo, 2014; Palmer, 1996). In contrast to directional asymmetry and anti-symmetry, fluctuating asymmetry does not favor one side; instead, it is random differences between the right and left sides of a bilateral trait or measurement (Graham, Raz, Hel-Or, & Nevo, 2010; Palmer, 1996; Van Valen, 1962). Franks and Cabo (2014: 500) describe this type of asymmetry as "small, random deviations from perfect symmetry that are not related to directional asymmetry or anti-symmetry and that, unlike the other two asymmetry types, is not adaptive." While it is generally accepted that directional asymmetry and anti-symmetry are related to genetic or environmental factors, several studies have attempted to link fluctuating asymmetry with environmental stress or genetic disturbances during the developmental period (e.g., Eriksen et al., 2017; Kieser, Groeneveld, & Silva, 1997; Kohn & Bennett, 1986; Møller, 2006; Özener, 2011). However, the causes of fluctuating asymmetry remain largely unconfirmed and controversial (see Graham et al., 2010; Graham & Özener, 2016).

Asymmetry in the human body has been amply documented, with topics including, but not limited to, facial asymmetry and attractiveness (e.g., Grammer & Thornhill, 1994; Hume & Montgomerie, 2000), brain laterality and language (e.g., Cantalupo & Hopkins, 2001), dermatoglyphic asymmetry and developmental stress (e.g., Arrieta et al., 1993; King, Dancause, Turcotte-Tremblay, Veru, & Laplante, 2012), soft tissue asymmetry and body weight (e.g., Domjanic, Fieder, Seidler, & Mitteroecker, 2013), and dental asymmetry and fitness (e.g., Bailit, Workman, Niswander, & Mac Lean, 1970) (see Graham & Özener, 2016 for a review). Skeletal asymmetries have also been well documented, with many studies reporting on the presence of directional asymmetries, especially regarding the limbs (e.g., Auerbach & Ruff, 2006; Cuk, Leben-Seljak, & Stefancic, 2001; Drapeau, 2008; Hiramoto, 1993; Kanchan, Kumar, Kumar, & Yoganarasimha, 2008; Latimer & Lowrance, 1965), and the relationship between directional asymmetries and handedness (see Ubelaker & Zarenko, 2012 for a review). Understanding directional asymmetries in the skeleton is particularly important, as notable morphological or metric differences between sides of an individual may complicate forensic and bioarchaeological analyses (Krishan & Kanchan, 2016). The following sections serve to illustrate this point.

Asymmetry from the lens of forensic anthropology and bioarchaeology

Potential impact of limb asymmetry on accurate sex estimation

Although limb asymmetry has been investigated using morphological features such as muscle insertion sites (e.g., Drapeau, 2008; Hawkey & Merbs, 1995), limb asymmetry research typically involves metric analysis of breadths, lengths, diameters, and weights. Several studies examining asymmetry in the limbs have shown that a contralateral relationship exists between upper body dominance and lower body dominance, where right-handed individuals exhibit right dominance of the upper limb (i.e., the right upper limb is longer/broader/heavier compared to the left) and left dominance of the lower limb (Auerbach & Ruff, 2006; Cuk et al., 2001; Kanchan et al., 2008; Krishan, 2011; Latimer & Lowrance, 1965). Although this "crossed symmetry" pattern has been observed in fetal remains prior to the influence of side preference (see Schultz, 1926), this pattern is generally thought to result from increased loading from bodyweight and muscle contractions that occur in the opposite side of the lower body during mechanical loading of the dominant upper limb (Auerbach & Ruff, 2006; Kanchan et al., 2008; Plochocki, 2002, 2004).

Although right-handed individuals are understood to exhibit right dominance of the upper limb and left dominance of the lower limb, the observed asymmetry in the lower body has been shown to be markedly reduced and more variable compared with asymmetry in the upper body (Auerbach & Ruff, 2006; Krishan, Kanchan, & DiMaggio, 2010). The bilateral differences in the degree of asymmetry observed in the upper and lower limbs can be attributed to differences in biomechanical loading of the upper and lower extremities. The dominant upper limb is subject to increased loading stress through preferential use, which leads to increased robusticity on the dominant side. Although the opposite lower limb will counter the increased loading from the dominant upper limb, the lower limbs are generally subject to more equal forces of magnitude that occur regularly during bipedal locomotion (Kanchan et al., 2008; Krishan et al., 2010). Therefore, while the lower left limb would be expected to be somewhat more robust than the lower right limb in right-handed individuals, the side difference in robusticity between the lower limbs is less drastic (see Auerbach & Ruff, 2006 for a discussion).

There remains a paucity of research regarding the impact of limb asymmetry on accurate sex classification despite limb asymmetry being a common phenomenon. Sex estimation methods that utilize long bone metrics, such as the widely used statistical software program FORDISC (Jantz & Ousley, 2005), rely on dimensions frequently shown to be asymmetric. For example, Auerbach and Ruff (2006) found significant levels of directional asymmetry in length measurements of the humerus, radius, and femur, as well as in breadth measurements of the humerus and femur, and also in diameter measurements of the humerus, radius, femur, and tibia. Therefore, understanding the impact of limb asymmetry on sex estimation is important. This is especially true considering that some breadth measurements that display directional asymmetry (e.g., humeral distal epicondylar breadth, femur epicondylar breadth) are considered the most important for sex estimation (e.g., Spradley & Jantz, 2011), and recent articles have shed light on the impact of limb asymmetry on other parameters, such as estimates concerning number of individuals and stature.

Kanchan et al. (2008) have discussed how the presence of directional asymmetry in long bone dimensions can cause confusion in estimating the number of individuals to whom remains belong, especially in mass disaster scenarios where elements are often numerous and comingled. These authors reported on the long bone dimensions of a single individual and found notable differences in length and weight between sides. Asymmetry was most pronounced in the upper limbs, with the right limb exhibiting larger dimensions compared with the left. Asymmetry was also present, but less pronounced in the lower limbs; and, as expected, based on the crossed symmetry pattern mentioned above, the contralateral limb (i.e., the left-sided elements) exhibited larger dimensions. Kanchan et al. (2008) warn that variations in long bone dimensions resulting from directional asymmetry may lead to an erroneous estimate of minimum number of individuals and recommend using additional corroboratory information before providing an estimate.

Krishan et al. (2010) investigated the impact of limb asymmetry on stature estimation by calculating six length dimensions of the extremities using a large sample ($n = 967$) of right-handed individuals. Five dimensions were found to exhibit significant levels of asymmetry and follow the crossed symmetry pattern. To test the impact of limb asymmetry on stature estimation, regression formulae were developed using these dimensions from both right and left sides, and measurements from the right side were subsequently tested on formulae developed using the left side. Findings indicate that stature estimates obtained using limb dimensions from the opposite side of the body from which the equations were developed result in erroneous estimates that are directly attributable to directional asymmetry (Krishan et al., 2010). Recently, Nandi, Olabiyi, Okubike, and Iheaza (2018) performed a similar study using upper limb lengths ($n = 230$) and when dimensions from the right limb were used with regression formulae derived using the left limb, significant differences in stature estimates were obtained. The authors of both studies strongly recommend using dimensions from the same side of the body from which stature equations have been developed to reduce the impact of asymmetry on method accuracy (Krishan et al., 2010; Nandi et al., 2018).

The abovementioned studies illustrate how limb asymmetry can complicate forensic and bioarchaeological analyses concerning number of individuals and stature, and previous research regarding the common occurrence of limb asymmetry strongly suggests that asymmetry in these regions could also impact sex estimates. Because limb asymmetry is a common phenomenon, and because postcranial metrics are important in sex estimation, especially in the absence of the pelvis (Spradley & Jantz, 2011), future research investigating the impact of asymmetry on sex estimates obtained from the long bones is a necessary endeavor.

Impact of skull and pelvic asymmetry on accurate morphological sex estimation

Several studies have investigated the occurrence of skull and pelvic asymmetry; however, they do not use the information for sex classification purposes, but rather focus on age estimation. Asymmetry in the timing and degree of cranial suture closure has been shown to be common, especially in the coronal and lambdoidal sutures, and assessing the younger side was found to provide more accurate results (Živanović, 1983). Asymmetry in the pubic symphysis has resulted in the assignment of different phases between sides when using the Suchey-Brooks method (Brooks & Suchey, 1990), with the older side being more accurate (Overbury, Cabo, Dirkmaat, & Symes, 2009). Research on phase- and component-based methods of the pubic symphysis and auricular surface has shown that information may be lost when using phase-based methods due to asymmetry in the progress of age-related traits between sides. It is recommended that component-based methods that incorporate information from both sides be used to obtain a single cohesive age estimate (McCormick & Kenyhercz, 2015). Additionally, a recent study using subadults has shown that epiphyses with shorter fusion periods (e.g., ischio-pubic ramus) have higher rates of asymmetry compared to epiphyses that fuse over many years (e.g., long bones). The accuracy of age estimation methods was impacted by asymmetry in epiphyseal fusion, and the authors suggest using all available elements rather than choosing the advanced/delayed or left/right side (Stull & Corron, 2017).

If asymmetry in the skull and pelvis can compromise age estimates, it is reasonable to predict that asymmetry in the skull and pelvis could compromise sex estimates. Recently, Cole (Cole, 2017; Cole, Cabo, & Klales, 2017) investigated the frequency, degree, and direction of asymmetry in the bilateral sex traits of the skull and pelvis ($n = 1818$; $F = 793$; $\mathrm{M} = 1025$) to determine the impact of asymmetry on accurate sex classification using the Walker (2008) and Klales, Ousley, and Vollner (2012) methods. The bilateral traits examined include the mastoid process (MP), supra-orbital margin (SO), ventral arc (VA), subpubic contour (SPC), and the medial aspect of the ischio-pubic ramus (MA). Neither the Walker (2008) nor Klales et al. (2012) methods recommend using a particular side for analysis, although Klales et al. (2012) indicate the left side was used in method creation. Although Walker (2008) does not indicate which side was used in method creation, illustrations provided to assist with method application depict the left side of the skull. In both methods, traits are scored on a scale of 1 (most gracile form) to 5 (most robust form).

Asymmetry was common in all bilateral traits examined and was more frequent in males for MP, SO, and VA, and more frequent in females for SPC and MA. The most frequently asymmetric trait overall was the MP (36.0% F, 41.0% M, 38.3% combined (C)). SO was the third most frequently asymmetric trait for both sexes (27.7% F, 32.2% M, 30.2% C). Interestingly, MA was the most frequently asymmetric pelvic trait for females (33.3%) and the least commonly asymmetric trait overall for males (23.4%; 27.6% C), while VA was the most frequently asymmetric pelvic trait for males (34.5%) and the least commonly asymmetric trait overall for females (21.6%; 29.1% C). Asymmetry in SPC was intermediate to the other two pelvic traits for both sexes and was the least asymmetric trait overall for combined sexes (25.6% F, 23.8% M, 24.6% C).

The degree of trait asymmetry was also examined (Table 1). The trait exhibiting the highest degree of asymmetry was SO. Despite being the most frequently asymmetric trait, the degree of asymmetry expressed in MP was relatively low. Conversely, SPC expressed the highest degree of asymmetry for the pelvic traits despite being the least frequently asymmetric trait for combined sexes.

Although asymmetries were common, the degree of asymmetry was relatively low, with most asymmetries consisting of only one score difference between sides. This observation may, in part, be explained as a result of intraobserver scoring inconsistencies in which many traits were scored by a difference of + or −1 score between trials. In Walls, Klales, Lesciotto, Gocha, and Garvin's (2018) analysis of intraobserver agreement using this data, Cohen's linear weighted kappa indicates substantial agreement between trials for these traits (0.66–0.89). However, weighted kappa does not provide a percent agreement between the two sides, which is essentially how asymmetry is quantified in Cole et al. (2017). When percent agreement was examined using the intraobserver error dataset from Walls et al. (2018), MP had the highest occurrence of disagreement between trials and hence the lowest weighted kappa (0.66). Score differences occurred in 77.1% of individuals between trials, but with only 11.1% of these being off by more than one. SO had the second highest occurrence of disagreement between trials (53.5% with only 6.1% off by more than one score) with a weighted kappa of 0.67 (Walls et al., 2018). Therefore, the high frequency of asymmetry for MP and SO in Cole et al. (2017) can largely be explained by intraobserver variation in trait scoring. Regarding the pelvic traits, MA had the highest occurrence of disagreement between trials (42.2%), followed by VA (38.2%) and SPC (31.4%). Although these percentages are somewhat high, the majority of inconsistencies between trials were of one score magnitude, and very few differed by more than one (VA = 3.9%, MA = 2.9%, SPC = 2.0%). These findings indicate that asymmetry in the bilateral traits of the Walker (2008) and Klales et al. (2012) methods is a real phenomenon; however, scoring inconsistency can result in inflated rates of asymmetry.

A chi-square analysis was used to test for directional asymmetries (Table 2). The skull traits exhibited right dominance (i.e., assigned the higher/more robust score) at $P < .05$ (Fig. 1).

Table 1 Degree of trait asymmetries by sex.

	Degree of asymmetry							
	1 score difference		2 score difference		3 score difference		4 score difference	
	Females	Males	Females	Males	Females	Males	Females	Males
MP	92.8%	86.1%	7.2%	13.0%	0%	0.7%	0%	0.2%
SO	76.9%	85.3%	19.3%	12.5%	2.4%	1.9%	1.4%	0.3%
VA	92.2%	86.8%	6.5%	10.5%	1.3%	2.4%	0%	0.3%
SPC	86.4%	87.8%	12.5%	10.5%	0.5%	1.7%	0.5%	0%
MA	90.0%	97.8%	9.6%	2.2%	0.4%	0%	0%	0%

Table 2 Direction of trait asymmetries by sex.

	Direction of asymmetry	
	Females	**Males**
MP	**Right**	**Right**
SO	**Right**	**Right**
VA	Right	Right
SPC	**Right**	**Right**
MA	Left	**Right**

Bold text indicates statistical significance at $P < .05$.
"Direction" refers to which element had the larger score.

Fig. 1 Asymmetry of MP exhibiting right dominance.

Fig. 2 Asymmetry of VA and SPC exhibiting right dominance.

The pelvic traits also generally exhibited right dominance at $P < .05$, although with less consistency (Fig. 2).

After establishing the frequency, degree, and direction of asymmetry in the bilateral traits of the Walker (2008) and Klales et al. (2012) methods, the accuracy of both methods was tested between symmetric and asymmetric individuals. Only five of the six equations provided by Walker (2008) were used because Walker's (2008) Equation 3 utilizes only unilateral traits of the skull and would, therefore, not be impacted by asymmetry. Because all three pelvic traits are used simultaneously in the Klales et al. (2012) method, it should be noted that individuals described as "symmetric" are symmetric in all three pelvic traits. Asymmetric individuals were grouped by individuals asymmetric in a single trait (VA, SPC, or MA only), two traits (VA/SPC, VA/MA, SPC/MA), and all three traits combined.

The presence of asymmetry was found to significantly decrease the classification accuracies of both methods depending on which traits were asymmetric and which equation was utilized (Tables 3 and 4). With regard to the Walker (2008) method, for all equations, asymmetric females classified better using the left side, and asymmetric males classified better using the right side. Using Klales et al. (2012), females were again more often correctly classified using the left side. Exceptions to this include females who were asymmetric in SPC only, MA only, and the combination of these two traits. In these situations, the rates of classification from the right and left sides were identical. Again, as was the case for the Walker (2008) method, asymmetric males generally classified better when the right side was used. Exceptions to this include males who were asymmetric in VA only and VA/MA combined. In the former case, the left side provided a slightly higher classification rate. In the latter case, the classification accuracies between sides were identical. The overwhelming occurrence of increased method accuracy for females when the left side was used—and for males when the right side was used—can be directly attributed to the directional asymmetry toward right dominance observed in most traits for both sexes.

Table 3 Classification accuracies using the Walker (2008) method for symmetric and asymmetric groups by sex.

Walker (2008)						
	Asymm females (*n* correct)	**Asymm females**	**Symm females**	**Asymm males (*n* correct)**	**Asymm males**	**Symm males**
L Equation 1	153/245	62.4%	63.5%	343/385	89.1%	93.0%
R Equation 1	143/245	58.4%		358/385	93.0%	
L Equation 2	164/276	59.4%	56.6%	367/409	89.7%	92.8%
R Equation 2	144/276	52.2%		381/409	93.2%	
L Equation 4	91/247	36.8%	37.2%	344/385	89.4%	94.1%
R Equation 4	79/247	32.0%		364/385	94.5%	
L Equation 5	50/190	26.3%	28.1%	286/305	93.8%	96.4%
R Equation 5	40/190	21.1%		292/305	95.7%	
L Equation 6	156/272	57.4%	55.0%	313/405	**77.3%**	91.3%
R Equation 6	121/272	44.5%		359/405	**88.6%**	

Bold text indicates statistical significance between symmetric and asymmetric groups at $P < .05$.

Table 4 Classification accuracies using the Klales et al. (2012) method for symmetric and asymmetric groups by sex.

Klales et al. (2012)						
	Asymm females (*n* correct)	Asymm females	Symm females	Asymm males (*n* correct)	Asymm males	Symm males
L VA	51/53	**96.2%**	98.5%	167/179	93.3%	95.5%
R VA	48/53	**90.6%**		165/179	92.2%	
L SPC	68/71	95.8%		79/83	95.2%	
R SPC	68/71	95.8%		80/83	96.4%	
L MA	136/139	97.8%		92/101	91.1%	
R MA	136/139	97.8%		94/101	93.1%	
LVA, SPC	37/40	**92.5%**		56/72	**77.8%**	
R VA, SPC	34/40	**85.0%**		59/72	**81.9%**	
L VA, MA	34/34	100%		45/49	91.8%	
R VA, MA	33/34	97.1%		45/49	91.8%	
L SPC, MA	38/39	97.4%		40/43	93.0%	
R SPC, MA	38/39	97.4%		41/43	95.3%	
L VA, SPC, MA	23/24	**95.8%**		25/31	**80.6%**	
R VA, SPC, MA	21/24	**87.5%**		30/31	**96.8%**	

Bold text indicates statistical significance between symmetric and asymmetric groups at $P < .05$.

The presence of asymmetry resulted in opposing sex estimates between sides in a number of individuals (Tables 5 and 6). Females rendered conflicting sex estimates more frequently than males for the Walker (2008) method with the reverse being true for the Klales et al. (2012) method. Although the number of individuals with opposing sex estimates is relatively small, it is important to be cognizant that the presence of directional asymmetry can nonetheless render conflicting estimates between the right and left sides of an individual.

Classification accuracies obtained for both symmetric and asymmetric individuals using the Walker (2008) method were well below the reported accuracies obtained by Walker (2008), and failed to reach the 85% threshold for acceptable adult sex estimation methods suggested by DiGangi and Moore (2013) (Tables 3 and 4). With few exceptions (e.g., Soficarua, Constantinescua, Culeaa, & Ionică, 2014), validation studies of Walker (2008) have also generally failed to achieve comparable classification accuracies (e.g., Garvin & Klales, 2018; Garvin, Sholts, & Mosca, 2014; Klales &

Table 5 Frequency of individuals rendering opposing sex estimates for the Walker (2008) method.

	Walker (2008)			
	Asymm females		**Asymm males**	
	n	**%**	***n***	**%**
Equation 1	54/245	22.0%	37/385	9.6%
Equation 2	80/276	29.0%	37/409	9.0%
Equation 4	72/247	29.1%	36/385	9.4%
Equation 5	23/190	12.1%	16/305	5.2%
Equation 6	133/272	48.9%	101/405	24.9%

Table 6 Frequency of individuals rendering opposing sex estimates for the Klales et al. (2012) method.

	Klales et al. (2012)			
	Asymm females		**Asymm males**	
	n	**%**	***n***	**%**
VA	7/53	13.2%	18/179	10.1%
SPC	0/71	0%	1/83	1.2%
MA	0/139	0%	2/101	2.0%
VA, SPC	3/40	7.5%	19/72	26.4%
VA, MA	1/34	2.9%	8/49	16.3%
SPC, MA	0/39	0%	1/43	2.3%
VA, SPC, MA	4/24	16.7%	7/31	22.6%

Cole, 2017; Krüger, L'Abbé, Stull, & Kenyhercz, 2015; Lewis & Garvin, 2016; Oikonomopoulou, Valakos, & Nikita, 2017).

In contrast to the Walker (2008) method, the Klales et al. (2012) method generally performed well for symmetric and asymmetric groups for both sexes. The symmetric groups for both sexes achieved accuracy rates in excess of 95%. Asymmetric females classified at rates at or above 85%. Asymmetric males classified at rates exceeding 91% for all trait combinations except when asymmetry affected both VA and SPC (both sides) and all three traits in combination (left side only). In these cases, the classification accuracy for asymmetric males was as low as 77.8% compared to the 95.5% accuracy achieved by symmetric males. This illustrates the impact that asymmetry can have on even the most accurate methods and why asymmetry is important to consider when performing analyses.

The detection of directional asymmetry toward right dominance in the Cole et al. (2017) study indicates that preferential selection of the left side causes a systemic decrease in the classification accuracy of males with the reverse being true if the right side is selected. Therefore, Cole et al. (2017) recommend assessing both sides of an individual whenever possible, and unless other regions of the skeleton or other methods are available to help inform a decision, a sex estimate should not be offered in situations where opposing sex estimates are obtained between right and left sides.

Further explorations of pelvic asymmetry

In an effort to better understand pelvic asymmetry, the authors sought to: (1) investigate the impact of sex, age, ancestry, region, and temporal period on the presence of morphological pelvic asymmetry, (2) examine the link between directional asymmetry in the pelvis and the sex biases observed in validation studies of the Phenice (1969) and Klales et al. (2012) methods, and (3) investigate the frequency and direction of asymmetry in pelvic measurements that are routinely used in sex estimation.

Impact of sex, age, ancestry, region, and temporal period on morphological pelvic asymmetry

The authors used a subset of the data used by Cole et al. (2017) ($n = 1063$) to explore a number of covariates to quantify their influence on the likelihood of asymmetry in the pelvis. The variables included sex (M, F), age (16–101 years), ancestry (black, white, Hispanic, Asian, Native American), region (United States, South Africa), and temporal period (historic, contemporary).

First, the authors subset the data into symmetric and asymmetric groups. To be considered asymmetric, an individual needed to exhibit at least one score difference between right and left sides for at least one trait. Using these groups, a new binary variable for symmetric/asymmetric was created, and the frequency of symmetric/asymmetric individuals was quantified for each covariate. Subsequently, the binary symmetric/asymmetric variables and other covariates were used in a logistic regression model: symmetric/asymmetric was the outcome variable, and the covariates were used as predictor variables. Only single-variable models were developed because of small sample sizes. The authors were most interested in the influence of these covariates on the prediction of asymmetry in the logistic regression models; therefore, odds ratios are reported to facilitate interpretation.

A total of 741 individuals were found to exhibit asymmetry in at least one pelvic trait, while 322 individuals were symmetric in all pelvic traits. When subset by indicator, percentages of asymmetry in the total sample were as follows: VA = 26.2% ($n = 279$), SPC = 19.6% ($n = 208$), and MA = 23.9% ($n = 254$). Because some of the predictor variables had vastly different sample sizes, and because logistic regression is highly susceptible to imbalanced classes, the authors mitigated the discrepancies in sample sizes by down-sampling. Down-sampling randomly samples from within the larger class so that the resulting class frequencies are equal. In an effort to ensure the random down-sampled

Table 7 Number of individuals from the original, unmodified sample that presented with at least one score difference between sides in at least one trait by time period.

	Sex	VA	SPC	MA
Contemporary	F	6	4	17
	M	23	9	16
Historic	F	86	83	117
	M	164	112	104

subset acted as a true representative of the larger sample, the sampling was iterated 60 times. A logistic regression was run for each iteration of sampling, and the presented numbers are averaged parameters.

Results indicate that temporal period has the greatest influence on the probability of exhibiting asymmetry, but region also matters. Historic individuals are 3.3 times more likely to express asymmetry (with a minimum of one score difference in at least one trait) compared to contemporary individuals, and individuals from the United States are three times more likely to express asymmetry than individuals from South Africa. The differential expression of asymmetry between contemporary and historic populations can be easily appreciated in Table 7.

The findings illustrate that two of the predictor variables are important in predicting asymmetry. While the authors do not encourage the use of logistic regression models to estimate if a partial skeleton may or may not exhibit asymmetry, the goal was to explore which demographic variables may influence the likelihood of presenting with asymmetric traits. The findings suggest that there may be differences in habitual tendencies among individuals from these different regions and temporal periods that result in differential percentages of the population that may present with asymmetry. An increased awareness of why asymmetry exists allows for a better understanding of how to contend with asymmetry in forensic anthropological and bioarchaeological contexts.

Link between directional asymmetry and sex biases observed in validation studies of the pelvis

Validation studies of the Phenice (1969) and Klales et al. (2012) methods have been conducted, but sometimes with seemingly contradictory sex bias results. To determine if the directional asymmetries detected in Cole et al. (2017) could explain the discrepancies in the reported sex biases of validation studies of the pelvis, the authors conducted a meta-analysis of the validation studies of the Phenice (1969) and Klales et al. (2012) methods. Unfortunately, not all studies indicate which side was used in the analysis, and several used both sides simultaneously; thus, a total of 10 validation studies could be examined.

A discrepancy in the reported sex bias was detected, with some studies reporting that the methods were more accurate when assessing females, and some studies reporting the methods were more accurate when assessing males. Where a side is specified, a female sex bias (i.e., higher classification accuracy) was obtained when the left side was used (e.g., Gómez-Valdés et al., 2017; Johnstone-Belford, Flavel, & Franklin, 2018; Kenyhercz, 2012; Kenyhercz, Klales, Stull, McCormick, & Cole, 2017; Klales et al., 2012; Klales & Cole, 2017; Lesciotto & Doershuk, 2018; Oikonomopoulou et al., 2017), and a male sex bias was obtained when the right side was used (e.g., Toon & Garcia de Leon, 2014). One exception to these findings is provided by MacLaughlin and Bruce (1990) who tested the Phenice (1969) method. These authors used the right side in performing their validation study and found a strong female sex bias. However, the results of their study have recently been called into question by McFadden and Oxenham (2016) who suggest that MacLaughlin and Bruce (1990) may have applied the method erroneously through their inclusion of an "ambiguous" sex category. Any individual assigned to the "ambiguous" category was considered to be misclassified, which is, McFadden and Oxenham (2016) note, a weighty assumption considering that a random assignment of "ambiguous" individuals would result in an accuracy rate around 50% instead of 0%. Recalculating the results obtained by MacLaughlin and Bruce (1990) for "ambiguous" individuals by assuming a 50% accuracy rate (random allocation), 80% accuracy rate (conservative estimate for experienced observers), and 96% accuracy rate (Phenice's (1969) reported accuracy) led to accuracy rates between 83% and 95% compared to the 59%–83% accuracy rates reported by MacLaughlin and Bruce (1990). The presence of morphological pelvic asymmetry and the directional trend toward right dominance detected in the Cole et al. (2017) study likely explain the sex bias phenomenon observed in these validation studies. These findings suggest that directional asymmetry should be examined and factored in all validation studies of methods concerning bilateral traits.

Frequency and direction of asymmetry in pelvic measurements

Literature involving metric pelvic asymmetry primarily pertains to clinical applications rather than sex estimation (e.g., Badii et al., 2003; Boulay et al., 2006). To investigate the frequency and direction of asymmetry in pelvic measurements routinely used for the estimation of sex, the authors analyzed eight measurements from a sample of 170 ($F=66$, $M=104$) contemporary white and black individuals (see Driscoll, 2010 for measurement definitions). To account for acceptable levels of interobserver error, individuals were considered asymmetric if the difference between the right and left sides was >2.0 mm.

Asymmetry was found to be common for both sexes, with frequencies ranging from 21.2% (minimum pubis length) to 62.9% (anterior superior iliac spine to symphysion) for combined sexes (Table 8). Asymmetry was found to be more frequent in males than

Table 8 Frequency and direction of asymmetry in pelvic measurements.

	Frequency			Direction	
	Females	Males	Combined sex	Females	Males
Max innominate height	40.9%	39.4%	40.0%	Left	Left
Max iliac breadth	42.4%	64.4%	55.9%	**Left**	**Left**
Max pubis length	21.2%	28.8%	25.9%	Left	Left
Min pubis length	25.8%	18.3%	21.2%	Left	Right
Min ischial length	24.2%	33.7%	30.0%	Right	Right
Anterior superior iliac spine to symphysion	54.5%	68.3%	62.9%	**Left**	**Left**
Max posterior superior iliac spine to symhysion	54.5%	62.5%	59.4%	**Left**	**Left**
Min apical border to symphysion	62.1%	55.8%	58.2%	**Right**	**Right**

Bold values indicate statistical significance at $P < .05$. "Direction" refers to which element had the larger measurement.

females for all measurements except maximum innominate height, minimum pubis length, and minimum apical border to symphysion (Table 8). When asymmetry was present, the larger measurement was most frequently observed on the left side (Table 8). It is important to note that the finding of left dominance (i.e., larger measurements) in this metric study is seemingly contradictory to the right dominance (i.e., greater robusticity) observed in the morphology study by Cole et al. (2017). However, because females generally have larger pelvic dimensions to facilitate the birthing process, "dominance" in the metric study refers to the more gracile form. Left dominance has been observed in other metric studies involving pelvic asymmetry (Badii et al., 2003; Tobolsky, Kurki, & Stock, 2016), and it has been speculated that this finding may be the result of crossed symmetry and the interaction and co-dependence of the pelvis with the lower limbs (Kurki, 2017; Tobolsky et al., 2016). Although the results obtained in the morphological pelvic study by Cole et al. (2017) do not follow the crossed symmetry pattern, it is speculated that the discrepancy between studies may result from differential use of the lower limbs, as well as the traits considered. The left dominance in pelvic measurements may be associated with mechanical loading of the stationary (contralateral) limb during manipulation tasks with the "preferred" lower limb, leading to right dominance of morphological pelvic traits, which serve as muscle attachment sites.

The information presented here illustrates that asymmetry in pelvic measurements is a common phenomenon and should be considered when employing metric techniques for sex estimation from this region, as the preferential selection of one side could cause a systematic decrease in the classification accuracy of one sex. The authors recommend that further research on this topic be conducted to better understand the full potential of metric pelvic asymmetry to impact sex estimation methods.

Conclusion

Skeletal asymmetries are present in a large proportion of the population. Previous research has demonstrated the impact of limb asymmetry on estimates concerning stature and calculating minimum number of individuals. Further, asymmetry in the skull and pelvis has been shown to impact age estimates. These findings inform the basis for exploring the potential impacts of skeletal asymmetry on other aspects of the biological profile. The implications of the information presented here demonstrate both the potential and reality of skeletal asymmetries to impact accurate sex classification.

Directional asymmetries were observed in the bilateral traits utilized by the Walker (2008) and Klales et al. (2012) methods, and despite the degree of trait asymmetries being low, the classification accuracies of both methods were shown to decrease when asymmetry was present. The directional trend toward right dominance for most traits led to improved classification accuracies for asymmetric females at the expense of decreased accuracies for asymmetric males when using the left side, with the reverse outcome when using the right side. Therefore, instead of preferentially selecting one side for analysis, Cole et al. (2017) recommend that both sides be evaluated when possible, and in cases where opposing sex estimates are obtained, a sex estimation should not be offered. A particularly interesting finding is that the presence of directional asymmetry likely explains the sex biases observed in validation studies of the Phenice (1969) and Klales et al. (2012) methods.

Practitioners should be mindful when estimating the sex of historic individuals in the United States, as the authors have demonstrated the increased frequency of morphological pelvic asymmetry in these populations. With regard to contemporary individuals, those from the United States presented with higher rates of pelvic asymmetry than those from South Africa. These findings have implications in both forensic anthropology and bioarchaeology, as understanding which individuals are more likely to present with asymmetry can help allow for more informed sex estimates to be made.

Finally, the authors investigated the frequency and direction of asymmetry in pelvic measurements. Asymmetry was found to be common, and a directional trend toward left dominance was detected. Future studies could investigate the potential impact of these findings on the classification accuracies of sex estimation methods that utilize pelvic measurements. To date, little research has been conducted on metric pelvic asymmetry, especially as it pertains to sex estimation, providing an important avenue for continued research on this topic. Thus, future research in this area is both a feasible and worthwhile endeavor.

Acknowledgments

The authors thank Kathryn R.D. Driscoll for contributing the metric pelvic data used in this chapter. Funding provided in part by the National Institute of Justice (2015-DN-BX-K014). Opinions or points of view expressed represent a consensus of the authors and do not necessarily represent the official position of the US Department of Justice.

References

Arrieta, M. I., Criado, B., Martinez, B., Lobato, M. N., Gil, A., & Lostao, C. M. (1993). Fluctuating dermatoglyphic asymmetry: Genetic and prenatal influences. *Annals of Human Biology*, *20*, 557–563.

Auerbach, B. M., & Ruff, C. B. (2006). Limb bone bilateral asymmetry: Variability and commonality among modern humans. *Journal of Human Evolution*, *50*, 203–218.

Badii, M., Shin, S., Torreggiani, W. C., Jankovic, B., Gustafson, P., Munk, P. L., et al. (2003). Pelvic bone asymmetry in 323 participants receiving abdominal CT scans. *Spine*, *28*, 1335–1339.

Bailit, H., Workman, P., Niswander, J., & Mac Lean, C. (1970). Dental asymmetry as an indicator of genetic and environmental conditions in human populations. *Human Biology*, *42*, 626–638.

Boulay, C., Tardieu, C., Benaim, C., Hecquet, J., Marty, C., Prat-Pradal, D., et al. (2006). Three-dimensional study of pelvic asymmetry on anatomical specimens and its clinical perspectives. *Journal of Anatomy*, *208*, 21–33.

Brooks, S. T., & Suchey, J. M. (1990). Skeletal age determination based on the os pubis: A comparison of the Acsadi-Nemeskeri and Suchey-Brooks methods. *Human Evolution*, *5*, 227–238.

Cantalupo, C., & Hopkins, W. D. (2001). Asymmetric Broca's area in great apes. *Nature*, *414*, 505.

Cole, S. J. (2017). *The effects of cranial and pelvic asymmetry on accurate nonmetric sex classification* [Master's thesis]. Erie, PA: Mercyhurst University.

Cole, S. J., Cabo, L. L., & Klales, A. R. (2017). The effects of cranial and pelvic asymmetry on accurate sex classification. In: *Proceedings of the 69th annual scientific meeting of the American Academy of Forensic Sciences; February 13–18; New Orleans, LA* (p. 118). Colorado Springs, CO: American Academy of Forensic Sciences.

Cuk, T., Leben-Seljak, P., & Stefancic, M. (2001). Lateral asymmetry of human long bones. *Variability and Evolution*, *9*, 19–23.

DiGangi, E. A., & Moore, M. K. (Eds.), (2013). *Research methods in human skeletal biology*. Oxford, UK: Elsevier Academic Press.

Domjanic, J., Fieder, M., Seidler, H., & Mitteroecker, P. (2013). Geometric morphometric footprint analysis of young women. *Journal of Foot and Ankle Research*, *6*, 1–8.

Drapeau, M. S. M. (2008). Enthesis bilateral asymmetry in humans and African apes. *Journal of Comparative Human Biology*, *59*, 93–109.

Driscoll, K. R. D. (2010). *Secular change of the modern human bony pelvis: Examining morphology in the United States using metrics and geometric morphometry* [Doctoral dissertation]. Knoxville, TN: University of Tennessee, Knoxville.

Eriksen, A. B., McGrath, K., Gómez-Robles, A., Schroeder, L., Massey, J. S., Bromage, T. G., et al. (2017). Facial fluctuating asymmetry in wild Virunga mountain gorillas (*Gorilla beringei beringei*). In *Proceedings of the 86th annual meeting of the American Association of Physical Anthropologists; Apr 19–22; New Orleans, LA* (p. 174): American Journal of Physical Anthropology.

Franks, E. M., & Cabo, L. L. (2014). Quantifying asymmetry: Ratios and alternatives. *American Journal of Physical Anthropology*, *154*, 498–511.

Garvin, H. M., & Klales, A. R. (2018). A validation study of the Langley et al. (2017) decision tree model for sex estimation. *Journal of Forensic Sciences*, *63*, 1243–1251.

Garvin, H. M., Sholts, S. B., & Mosca, L. A. (2014). Sexual dimorphism in human cranial trait scores: Effects of population, age, and body size. *American Journal of Physical Anthropology*, *154*, 259–269.

Gómez-Valdés, J. A., Garmendia, A. M., García-Barzola, L., Sánchez-Mejorada, G., Karam, C., Baraybar, J. P., et al. (2017). Recalibration of the Klales et al. (2012) method of sexing the human innominate for Mexican populations. *American Journal of Physical Anthropology*, *162*, 600–604.

Graham, J. H., & Özener, B. (2016). Fluctuating asymmetry of human populations: A review. *Symmetry*, *8*, 1–36.

Graham, J. H., Raz, S., Hel-Or, H., & Nevo, E. (2010). Fluctuating asymmetry: Methods, theory, and applications. *Symmetry*, *2*, 466–540.

Grammer, K., & Thornhill, R. (1994). Human (*Homo sapiens*) facial attractiveness and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology*, *108*, 233–242.

Hawkey, D. E., & Merbs, C. F. (1995). Activity induced musculoskeletal stress makers (MSM) and subsistence strategy changes among ancient Hudson Bay Eskimos. *International Journal of Osteoarchaeology*, *5*, 324–338.

Hiramoto, Y. (1993). Right-left differences in the lengths of human arm and leg bones. *Kaibogaku Zasshi*, *68*, 536–543.

Hume, D. K., & Montgomerie, R. (2000). Facial attractiveness signals different aspects of "quality" in women and men. *Evolution and Human Behavior*, *22*, 93–112.

Jantz, R. L., & Ousley, S. D. (2005). *FORDISC 3: Computerized forensic discriminant functions*. Version 3.1 Knoxville, TN: The University of Tennessee.

Johnstone-Belford, E., Flavel, A., & Franklin, D. (2018). Morphoscopic observations in clinical pelvic MDCT scans: Assessing the accuracy of the Phenice traits for sex estimation in a Western Australian population. *Journal of Forensic Radiology and Imaging*, *12*, 5–10.

Kanchan, T., Kumar, T. S. M., Kumar, G. P., & Yoganarasimha, K. (2008). Skeletal asymmetry. *Journal of Forensic and Legal Medicine*, *15*, 177–179.

Kenyhercz, M. W. (2012). Sex estimation using pubic bone morphology in a modern South African sample: a test of the Klales et al. method. In *Proceedings of the 81st annual meeting of the American Association of Physical Anthropologists; April 11–14; Portland, OR* (pp. 179–180): American Journal of Physical Anthropology.

Kenyhercz, M. W., Klales, A. R., Stull, K. E., McCormick, K. A., & Cole, S. J. (2017). Worldwide population variation in pelvic sexual dimorphism: A validation and recalibration of the Klales et al. method. *Forensic Science International*, *277*, 259.e1–259.e8.

Kieser, J. A., Groeneveld, H. T., & Silva, D. (1997). Dental asymmetry, maternal obesity, and smoking. *American Journal of Physical Anthropology*, *102*, 133–139.

King, S., Dancause, K., Turcotte-Tremblay, A. M., Veru, F., & Laplante, D. P. (2012). Using natural disasters to study the effects of prenatal maternal stress on child health and development. *Birth Defects Research Part C*, *96*, 273–288.

Klales, A. R., & Cole, S. J. (2017). Improving nonmetric sex classification for Hispanic individuals. *Journal of Forensic Science*, *62*, 975–980.

Klales, A. R., Ousley, S. D., & Vollner, J. M. (2012). A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *American Journal of Physical Anthropology*, *149*, 104–114.

Kohn, L. A., & Bennett, K. A. (1986). Fluctuating asymmetry in fetuses of diabetic rhesus macaques. *American Journal of Physical Anthropology*, *71*, 477–483.

Krishan, K. (2011). Marked limb bilateral asymmetry in an agricultural endogamous population of North India. *American Journal of Human Biology*, *23*, 674–685.

Krishan, K., & Kanchan, T. (2016). Forensic and anthropological application of body asymmetry: A comment on Gutnik et al. (2015). *Perceptual and Motor Skills*, *122*, 578–582.

Krishan, K., Kanchan, T., & DiMaggio, J. A. (2010). A study of limb asymmetry and its effects on estimation of stature in forensic casework. *Forensic Science International*, *200*, 181.e1–181.e5.

Krüger, G. C., L'Abbé, E. N., Stull, K. E., & Kenyhercz, M. W. (2015). Sexual dimorphism in cranial morphology among modern South Africans. *International Journal of Legal Medicine*, *129*, 869–875.

Kurki, H. K. (2017). Bilateral asymmetry in the human pelvis. *The Anatomical Record*, *300*, 653–665.

Latimer, H. B., & Lowrance, E. W. (1965). Bilateral asymmetry in weight and in length of human bones. *The Anatomical Record*, *152*(2), 217–224.

Lesciotto, K. M., & Doershuk, L. J. (2018). Accuracy and reliability of the Klales et al. (2012) morphoscopic pelvic sexing method. *Journal of Forensic Science*, *63*, 214–220.

Lewis, C. J., & Garvin, H. M. (2016). Reliability of the Walker cranial nonmetric method and implications for sex estimation. *Journal of Forensic Science*, *61*, 743–751.

MacLaughlin, S. M., & Bruce, M. F. (1990). The accuracy of sex identification in European skeletal remains using the Phenice characteristics. *Journal of Forensic Science*, *35*, 1384–1392.

McCormick, K., & Kenyhercz, M. (2015). Bilateral asymmetry in age-related trait expression of the pubic symphysis and auricular surface of the ilium. In *Proceedings of the 67th annual scientific meeting of the American Academy of Forensic Sciences; February 16–21; Orlando, FL* (p. 78). Colorado Springs, CO: American Academy of Forensic Sciences.

McFadden, C., & Oxenham, M. F. (2016). Revisiting the Phenice technique sex classification results reported by MacLaughlin and Bruce (1990). *American Journal of Physical Anthropology*, *159*, 182–183.

Møller, A. P. (2006). A review of developmental instability, parasitism and disease. *Infection, Genetics and Evolution*, *6*, 133–140.

Nandi, M. E., Olabiyi, O. A., Okubike, E. A., & Iheaza, E. C. (2018). A study of bilateral asymmetry of upper extremities and its effects on stature reconstruction amongst Nigerians. *Arab Journal of Forensic Sciences and Forensic Medicine*, *1*, 978–988.

Oikonomopoulou, E. K., Valakos, E., & Nikita, E. (2017). Population-specificity of sexual dimorphism in cranial and pelvic traits: Evaluation of existing and proposal of new functions for sex assessment in a Greek assemblage. *International Journal of Legal Medicine*, *131*, 1731–1738.

Overbury, R. S., Cabo, L. L., Dirkmaat, D. C., & Symes, S. A. (2009). Asymmetry of the os pubis: Implications for the Suchey-Brooks method. *American Journal of Physical Anthropology*, *139*, 261–268.

Özener, B. (2011). Does urban poverty increase body fluctuating asymmetry? *Collegium Antropologicum*, *35*, 1001–1005.

Palmer, A., & Strobeck, C. (1986). Fluctuating asymmetry: Measurement, analysis, patterns. *Annual Review of Ecology and Systematics*, *17*, 391–421.

Palmer, A. R. (1996). Waltzing with asymmetry. *Bioscience*, *46*, 518–532.

Phenice, T. W. (1969). A newly developed visual method of sexing the os pubis. *American Journal of Physical Anthropology*, *30*, 297–301.

Plochocki, J. H. (2002). Directional bilateral asymmetry in human sacral morphology. *International Journal of Osteoarchaeology*, *12*, 349–355.

Plochocki, J. H. (2004). Bilateral variation in limb articular surface dimensions. *American Journal of Human Biology*, *16*, 328–333.

Schultz, A. H. (1926). Fetal growth of man and other primates. *The Quarterly Review of Biology*, *1*, 465–521.

Soficarua, A., Constantinescua, M., Culeaa, M., & Ionică, C. (2014). Evaluation of discriminant functions for sexing skulls from visually assessed traits applied in the Rainer Osteological collection (Bucharest, Romania). *Journal of Comparative Human Biology*, *65*, 464–475.

Spradley, M. K., & Jantz, R. L. (2011). Sex estimation in forensic anthropology: Skull versus postcranial elements. *Journal of Forensic Science*, *56*, 289–296.

Stull, K. E., & Corron, L. (2017). The effects of epiphyseal fusion asymmetry on juvenile age estimation. In *Proceedings of the 86th annual meeting of the American Association of Physical Anthropologists; April 19–22; New Orleans, LA* (p. 373): American Journal of Physical Anthropology.

Tobolsky, V. A., Kurki, H. N., & Stock, J. T. (2016). Patterns and directional asymmetry in the pelvis and pelvic canal. *American Journal of Human Biology*, *28*, 804–810.

Toon, C., & Garcia de Leon, J. (2014). A comparison of the Klales et al. (2012) and Phenice (1969) methods of sex estimation on a modern Colombian sample. In *Proceedings of the 64th annual scientific meeting of the American Academy of Forensic Sciences; Feb 17–22; Seattle, WA* (p. 465). Colorado Springs, CO: American Academy of Forensic Sciences.

Ubelaker, D. H., & Zarenko, K. M. (2012). Can handedness be determined from skeletal remains? A chronological review of the literature. *Journal of Forensic Sciences*, *57*, 1421–1426.

Van Valen, L. (1962). A study of fluctuating asymmetry. *Evolution*, *16*, 125–142.

Walker, P. L. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, *136*, 39–50.

Walls, M., Klales, A. R., Lesciotto, K. M., Gocha, T. P., & Garvin, H. M. (2018). Reliability and validity of the Walker and Klales et al. methods. In *Proceedings of the 70th annual scientific meeting of the American Academy of Forensic Sciences; February 19–24; Seattle, WA* (p. 43). Colorado Springs, CO: American Academy of Forensic Sciences.

Živanović, S. (1983). A note on the effect of asymmetry on suture closure in mature human skulls. *American Journal of Physical Anthropology*, *60*, 431–435.

CHAPTER 20

Cognitive bias in sex estimation: The influence of context on forensic decision-making

Sherry Nakhaeizadeh, Itiel E. Dror, Ruth M. Morgan
UCL Department of Security and Crime Science, Centre for the Forensic Sciences, London, United Kingdom

Introduction

Forensic science has played an important role in criminal investigations and the legal process for centuries (Found, 2015). In recent years however, concerns about the lack of rigorous scientific research within the forensic science domains have been expressed in published literature and key governmental reports (e.g., the UK Government Chief Scientific Adviser, 2015; the US National Academy of Science, 2009; the US President's Council of Advisors on Science and Technology, 2016). Indeed, the effect of cognitive bias on the interpretation of forensic science evidence has been identified as a pressing issue across numerous forensic domains (e.g., Warren, Friend, & Stock, 2018). The impact of cognitive biases is being addressed at each stage of the forensic science process, including data collection, analysis, evidence interpretation, and final presentation in court (Edmond et al., 2017; Found, 2015; Morgan et al., 2019). It has been demonstrated that these vulnerabilities are not limited to a specific field, with similar cognitive biasing issues being observed empirically across many forensic science domains (e.g., Dror, Charlton, & Péron, 2006; Dror & Hampikian, 2011; Kukucka & Kassin, 2014; Laber et al., 2014; Miller, 1987, 1984; Nakhaeizadeh, Dror, & Morgan, 2014; Nakhaeizadeh, Hanson, & Dozzi, 2014; Nakhaeizadeh, Morgan, Rando, & Dror, 2018; Stevenage & Bennett, 2017) including forensic anthropology (Nakhaeizadeh & Morgan, 2015).

Within forensic anthropology, there has been critique of some of the techniques used by forensic anthropologists (e.g., Christensen & Crowder, 2009). Discussion has been extensive concerning evidence validation, admissibility, and error rates in the methods applied (Christensen, 2004; Christensen & Crowder, 2009). For example, it has been argued that some of the methods used in forensic anthropology are generally reliant upon observations and the specific experience of the analyst (Byers, 2010; Cattaneo et al., 1999; Dirkmaat, Cabo, Ousley, & Symes, 2008; Hefner, 2009). Some have contested some of the techniques, asserting that they are limited because of their subjective nature

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00020-1

(Lottering, MacGregor, Meredith, Alston, & Gregory, 2013; Spradley & Jantz, 2011; Walrath, Turner, & Bruzek, 2004) in a manner akin to other forensic disciplines. In response, there has been some modification of existing methods that have accompanied the development of new comparative samples in forensic anthropology and statistical tools for data analysis (Clark, Guatelli-Steinberg, Hubbe, & Stout, 2016; Dirkmaat & Cabo, 2012; Grivas & Komar, 2008; Hefner & Ousley, 2014; Langley, Dudzik, & Cloutier, 2018; Mahakkanukrauh, Ruengdit, Tun, Case, & Sinthubua, 2017; Spradley & Jantz, 2011; Walker, 2008). These developments have enhanced the role of quantitative methods and led to a rise in new publications in the literature concerning the analysis of skeletal remains, especially pertaining to sex estimation (e.g., Klales, Ousley, & Vollner, 2012).

However, the presence of cognitive bias, its impact, and the cognitive processes involved in the assessment of human remains have only recently begun to be assessed (Klales & Lesciotto, 2016; Nakhaeizadeh, Dror, & Morgan, 2014; Nakhaeizadeh, Hanson, & Dozzi, 2014). This chapter will focus on the effect of cognitive bias in forensic anthropology, with a particular focus on cognitive bias research within sex estimation. The chapter will include a brief introduction to the role of human cognition in decision-making, highlighting how these could affect expert performance, drawing on previous and current research within psychology and forensic science. The chapter will also underline the potential effect of cognitive bias in forensic anthropology, illustrating how experts might be prone to cognitive interpretation issues, by referencing current empirical research. A discussion of possible reforms—and then recommendations for future directions to explore and better develop our understanding of cognitive bias and minimize its impact in the practice of forensic anthropology broadly, and sex estimation specifically—is offered.

Human cognition and cognitive bias

Information processing in decision-making is part of human cognition and defines the acquisition, organization, and the use of knowledge (Anderson, 2000; Bandura, 1999; Wyer & Srull, 1986). The study of human cognition examines, among others, human perception, judgment, and decision-making, which are all influenced by a variety of cognitive processes (Hoppitt, Mathews, Yiend, & Mackintosh, 2010). Decades of research within psychology and human cognition have shown that the human mind relies heavily on its prior experiences, beliefs, emotions, and knowledge (top-down information) when encoding information. This system allows the brain to create strategies and "mental shortcuts" to help make sense of the information and data coming in (bottom-up), allowing for its decision-making to be quicker, more prudent, and accurate (Elstein, 1999; Gigerenzer & Gaissmaier, 2011).

However, emotions, prior experiences, prior knowledge, and prior beliefs can sometimes interfere and distort information processing, especially when we are making decisions under uncertainty. This can result in a decision-maker being susceptible to these contextual influences when interpreting the meaning of evidence (Dror, 2011; Girotto & Politzer, 1990). These biasing affects can be referred to as cognitive biases, generally defined as the psychological and cognitive factors that unconsciously manipulate and interfere with data processing, causing judgment and decision-making to be unreliable (Evans & Pollard, 1990).

The vast body of literature within psychology has, over the years, distinguished between different sources of cognitive bias, such as time pressure (Ordonez & Benson III, 1997), expectations (Bressan & Dal Martello, 2002), pre-existing beliefs (Hamilton & Zanna, 1974), and motivation (Kunda, 1990). For example, a series of studies by Balcetis and Dunning (2006) showed that the impact of motivation on information processing led participants to perceive visual stimuli that they desired. Moreover, the studies demonstrated that participants tended to interpret an ambiguous figure in a manner that "fitted" their preference and wishes. The research in this area revealed that perception is selective and malleable and highly related to the context within which the decision is being made. For example, the understanding of how "steep" a hill might be will be more extreme if participants are asked to make that estimation after they have jogged for an hour (Bhalla & Proffitt, 1999). Similarly, an estimation of the speed of a person will be biased if participants are asked to make that estimation after viewing very fast animals (such as a cheetah) versus very slow animals (such as a turtle) (Aarts & Dijksterhuis, 2002).

Equally, research has also shown that prior expectations could provide a sufficient and unconscious tendency to perceive and interpret evidence that would confirm preexisting beliefs. This is also otherwise known as confirmation bias (Khaneman & Frederick, 2002), which is the tendency to selectively gather and process information to confirm a hypothesis or preconception (Dror & Charlton, 2006; Gianelli, 2007). Over the years, confirmation bias has come to provide an umbrella term for a number of distinct ways that expectations and beliefs influence memory, selection, and evaluation of evidence (Nickerson, 1998), which has also been studied with regard to the role of expertise in decision-making.

Experts have specific cognitive mechanisms that are needed to perform certain tasks associated within their expert domains (Dror, 2016). For example, "expertise" of an expert can be acquired by repeated exposure to the tasks they need to perform, creating schemas from learning and experiences (Dror, 2011; Morgan, 2017b). Indeed, experts' reliance on top-down information allows for enhanced, quicker, and efficient performance, learning how to "automatically" filter out noise and deal with large amount of information (Edmond et al., 2017; Stanovich, 2014). This leads to experts being able to perform skills relatively effortlessly. Paradoxically, the cognitive architecture involved in being an expert has also been argued to result, in some situations, in a lack of flexibility, resulting in experts missing or ignoring important information (Sternberg, 2008). Arguably, the specialized nature of expertise can also render experts inflexible, and be especially prone to external influences (Edmond et al., 2017). In addition, numerous studies across different fields have shown that experience and exposure to a procedure does necessarily translate into expertise (Edmond et al., 2017) with, for example, studies in clinical psychology showing that a clinical psychologist's professional experience and length of training is not related to treatment success and efficiency (Dawes, 1994).

The growing concerns over expert decision-making being influenced by cognitive processes have led to a rise in research specifically focusing on applying different judgment and decision-making theories within forensic interpretations (Edmond et al., 2017; Found, 2015). This has led to research within expertise, decision-making, and situation awareness literature shifting its focus to not only concern human judgments in the social, psychological, and behavioral economics domains but also more recently within law enforcement agencies and forensic disciplines (Archer & Wallman, 2016; Ask & Granhag, 2005; Biedermann, Bozza, & Taroni, 2016; Dror & Charlton, 2006; Dror & Hampikian, 2011; Earwaker, Morgan, Harris, & Hall, 2015; Kerstholt et al., 2010; Mattijssen, Kerkhoff, Berger, Dror, & Stoel, 2016; Nakhaeizadeh et al., 2018; Osborne, Taylor, & Zajac, 2016).

Research in cognitive bias and forensic science

Within forensic science, research has begun to empirically address how cognitive bias can influence a wide range of forensic judgments. Many fields of forensic science include subjective assessment and comparison stages that are potentially susceptible to cognitive bias, due to their heavy reliance on human judgments (Kassin, Dror, & Kukucka, 2013; Thompson & Cole, 2007). Studies about cognitive bias in forensic science have shown that situational context, early hypothesis, and expectations can influence how evidence is collected, perceived, and interpreted.

For example, social interaction, past experiences, and prior information has been argued to influence forensic handwriting and document examinations in their final conclusions (Kukucka & Kassin, 2014; Miller, 1984; Stoel, Dror, & Miller, 2014). In addition, the effect of contextual information has also been shown within fingerprint examiners with regard to whether or not two fingerprint marks originate from the same source (Dror et al., 2006, 2011; Dror & Charlton, 2006). In many of these experiments, the majority of experts reached different conclusions on previously assessed fingermark comparison, revealing an inconsistency in their analysis when provided with new contextual information (Dror & Charlton, 2006; Dror, Peron, Hind, & Charlton, 2005).

Furthermore, in some studies, researchers point out that some stimuli are based on perceptual judgments that can cause a lack of interrater and intra-test consistency.

This has been shown within fingermark comparisons where findings show that, even without the context of the comparison print, there was still a lack of consistency in analyzing some latent marks (Dror et al., 2011; Langenburg, Champod, & Wertheim, 2009; Schiffer & Champod, 2007). Not only was this reflected by inconsistency between different experts, but also the same experts at different times were inconsistent with their own analysis (Dror et al., 2011).

Equally, the effect of context and potential for confirmation bias has also been identified within DNA (Dror & Hampikian, 2011), bite mark comparisons (Osborne, Woods, Kieser, & Zajac, 2014; Page, Taylor, & Blenkin, 2012), bloodstain analysis (Osborne et al., 2016), forensic entomology (Archer & Wallman, 2016), and fire scene examinations (Bieber, 2012). For example, the interpretation of a mixed DNA sample differed among DNA experts depending on the case context (Dror & Hampikian, 2011).

The findings within cognitive and contextual bias in forensic science have shown that expertise does not prevent the effect of context on decision-making (Dror, 2011; Edmond et al., 2017; Found & Ganas, 2013; van den Eeden, de Poot, & Van Koppen, 2016). A study that addressed cognitive bias in crime scene investigation showed that experts interpreted the crime scene differently depending on the prior information that the examiners were exposed to (van den Eeden, de Poot, & van Koppen, 2018). In fact, the study demonstrated that experienced crime scene investigators were not immune to the bias and were impacted as much as novices.

Despite the results of these published studies, alongside decades of research within psychology and social sciences, it has been argued that many examiners still have only a limited appreciation of cognitive bias and its impact within their own discipline (Kukucka, Kassin, Zapf, & Dror, 2017). A recent global survey on forensic examiners, and their beliefs about the scope and nature of cognitive bias, showed a "bias blind spot," where some forensic science examiners recognized that cognitive bias is a problem in forensic science, but denied that these biases could affect them personally (Kukucka et al., 2017). This might indicate that the empirical evidence base that underpins how individuals makes decisions, what influences those decisions, and how to enhance decision-making outcomes is still not fully appreciated in all forensic domains, including forensic anthropology (Nakhaeizadeh et al., 2018).

Cognitive bias and forensic anthropology

In forensic anthropology, the cognitive impacts in play during the assessment of human remains have only recently begun to be assessed in the published literature (Klales & Lesciotto, 2016; Nakhaeizadeh, Dror, & Morgan, 2014; Nakhaeizadeh, Hanson, & Dozzi, 2014). Like most forensic disciplines, human observation and qualitative opinion-based methods are commonly used in forensic anthropology to make interpretations (Christensen & Crowder, 2009; Grivas & Komar, 2008). Therefore, there is a level of subjectivity inherent to forensic anthropological practice (Warren et al., 2018). However, this does not necessarily mean that the methods are unreliable or less valid, but rather that they are generally reliant upon human observation and the specialized experience of the observer (Dirkmaat et al., 2008; Hefner, 2009). Nevertheless, the subjectivity inherent in some of the methods employed, in combination with the context to which a forensic anthropologist may be exposed (such as background stories, historical context, prior knowledge about the case details, etc.), can sometimes create conditions in which interpretations might be affected by cognitive influences.

Sex assessments and cognitive bias

The first step when generating a biological profile of an unidentified individual is the estimation of sex (Guyomarc'h & Bruzek, 2011). This is primarily due to some of the traditional methods applied for age estimation, ancestry, and stature being sex-specific (Klales, 2013). For example, the observable differences in aging and growth patterns between sexes, as well as variations in morphological traits related to ancestry, make accuracy of sex estimations vital (Krishan et al., 2016). Many have argued that the accuracy of sexing skeletal remains greatly depends on the element present for analysis and its preservation state (Đurić, Rakočević, & Đonić, 2005; Naikmasur, Shrivastava, & Mutalik, 2010), as well as the experience of the observer. Historically, the pelvic bone has been argued to be the most reliable single bone for sex estimation.

Some of the most extensively adopted sexing techniques are based on morphological observations and rely on the visual assessment of sexual dimorphic traits (Mahfouz et al., 2007). These assessments are generally used by forensic anthropologists due to their efficiency, as well as their practicality (Biwasaka et al., 2009; Đurić et al., 2005). However, the methods used in sexual dimorphic traits have been argued to be influenced by their level of subjectivity (Kemkes-Grottenthaler, Löbig, & Stock, 2002; Steyn, Pretorius, & Hutten, 2004). In addition, visual assessments in sex estimations generally show higher accuracy results with intact bones, with the level of accuracy tending to decrease with incomplete and fragmented skeletons (Krishan et al., 2016; Thomas, Parks, & Richard, 2016).

Metric assessments have been acknowledged to enable easier application of quantitative statistical analyses with associated error and probabilistic estimations (Kimmerle, Ross, & Slice, 2008; Spradley & Jantz, 2011). However, the accuracy in sexing based on metric assessments may vary significantly depending on the statistical model utilized (Krishan et al., 2016). Further, it is difficult to attribute differences in size to sex without considering ecological and physiological implications (Garvin, 2012). Not only are metric analyses limited, owing to issues of variation in size within pertinent populations, but this form of assessment also requires intact skeletal elements. It is worth noting, however, that while traditional metric methods are objective in essence, they often also suffer from observer discrepancies if landmarks are not properly defined (Krishan et al., 2016).

One of the earliest studies into the possibility of "biases" in sex assessments was conducted by Weiss (1972) in his work on "systematic bias in skeletal sexing." In this study, Weiss compared samples from archaeological skeletal populations and the accuracy of sex estimations on the skeletal sample. The results demonstrated a "male bias" in the assessment of skeletal remains, with 12% more males than expected when compared to sex ratios in living populations. Weiss argued (after a further analysis of the dataset) that the flaws in sexing methodologies were more likely to be compounded by bias rather than the population actually containing more males than females. He concluded that this bias was due to the nature of secondary sex characteristics in bones, and that this was particularly notable when assessing robust ambiguous skulls, as he argued that there is a tendency to misidentify ambiguous "robust" female skulls as males (Weiss, 1972). Weiss contended that this might be due to subtle societal prejudices in the field with regard to robust skeletal skulls being "expected" to be male morphological traits, perhaps resulting in a default male classification. Weiss also hypothesized that the general nature of sex characteristics on bones, in many cases, produced an "irresistible" call to classify doubtful specimens as male. Weiss argued that this could be due to the fact that greater weight is put on the sexing of specimens by characteristics of the skull, and that traits that are found to be of intermediate size seem to be more often classified as male traits rather than undetermined.

Walker (1995) further highlighted that there might be a societal prejudice of male and female characteristics (i.e., females appearing more gracile, and males more robust) that could potentially bias the interpretation of archaeological skeletal collections (Walker, 1995). For example, Walker's (1995) study of the well-documented Saint Brides Church skeletal collection in London showed that poorly preserved female pelvises with robust skulls were often misclassified as males. Walker (1995) noticed that female skulls in the studied population became more robust with age. Similar to Weiss, Walker hypothesized that this accounts for the prevalent misidentification of elderly females as males.

Furthermore, some have identified a potential for expectation and context bias issues in sex estimations of skeletal remains, where grave artifacts could potentially cause an expectancy bias in the interpretation due to the associated grave artifacts (Effros, 2000). In one recent study, Hedenstierna-Jonson et al. (2017) correctly identified (through DNA analysis) a Viking warrior—who has long been regarded as a male—to actually be a female. The study highlighted that the grave artifacts associated with the Viking warrior burial had a strong contextual (and arguably led to a confirmation bias affect) in the sex assessment of skeletal remains. Although the osteological analysis initially indicated the skeletal remains to be of a female, for over a century the Viking warrior was misidentified as a male due to the grave artifacts associated with the burial as well as the stereotypes of male Viking warriors (Hedenstierna-Jonson et al., 2017).

Empirical studies that have specifically addressed contextual affects and confirmation bias in sex estimations have shown that there can be an impact on the interpretations made by participants (Klales & Lesciotto, 2016; Nakhaeizadeh, Dror, & Morgan, 2014). One study by Nakhaeizadeh, Dror, and Morgan (2014) involved examining the effect of context on morphological estimations in sex, ancestry, and age-at-death. In the study, participants were asked to create a biological profile on ambiguous skeletal remains from one individual. Participants in the study were semi-randomly assigned into one of three groups, where two of the groups were given contextual information before conducting the analysis, with a third group acting as a control with no context provided. The contextual information was provided before establishing a biological profile and included context that gave indications of sex, origin, and age of the remains (e.g., giving participants the result of a DNA analysis of the remains). Similar to previous studies in contextual and confirmation bias in forensic science (e.g., Dror & Charlton, 2006; Dror & Hampikian, 2011), the study sought to determine if the examiners would be affected by the given context when asked to establish a biological profile. The results showed a difference in the interpretation of skeletal remains between participants depending on the context. For example, in the group that received contextual information prior to the analysis that the remains were female, 100% of the participants concluded the remains to be female. However, in the group that received contextual information that the remains were male, only 14% indicated the remains to be female, 72% indicated the skeletal remains to be male, and 14% were undetermined in their conclusion (Nakhaeizadeh, Dror, & Morgan, 2014).

Another study addressing confirmation bias in forensic anthropology was conducted by Klales and Lesciotto (2016). Here, the authors explored the idea of confirmation bias and sex estimations of the innominate. The study was conducted on 15 innominates with seven experienced observers asked to blindly score the three main traits outlined by Phenice (1969) within the context of the Klales et al. (2012) methodology. This was done using a developed five-scale scoring system, with 1 being gracile expression and 5 being robust expression. Each of the three traits was scored on a separate day with only the specific trait under examination being visible. After assessing each trait individually, participants were asked to provide an overall impression of the sex as well as scoring each trait again. However, this time all traits were visible and scored simultaneously in combination with examining the whole innominate. The results showed a tendency to change the scaling of single traits on the innominate that have been assessed in isolation to fit the overall decision reached, indicating a confirmation bias (Klales & Lesciotto, 2016).

A further study on early exposure to information at the crime scene was shown to have a subsequent effect on sex assessments subsequently made in the laboratory (Nakhaeizadeh et al., 2018). In this study, participants investigated a mock crime scene, which included the excavation of clandestine burials that had male skeletal casts dressed either in female or gender-neutral clothing, followed by a forensic anthropological assessment of the skeletal remains. The results indicated that the interpretation of sex estimation was highly dependent upon the context in which participants were exposed to prior to the analysis. For example, a high proportion of participants who were exposed to the female clothing context interpreted the skeletal remains to be female with only one participant determining the male skeletal cast to be male, showing a potential for a *cascading bias effect* (Dror, Morgan, Rando, & Nakhaeizadeh, 2017).

Although these studies all had inherent limitations, they still highlight that, under certain conditions, cognitive factors can influence the decision-making process, ultimately affecting the interpretation of skeletal remains. As mentioned previously, previous validation and classification studies within methods used in forensic anthropology have generally shown the methods to be reliable, with high classification accuracy, specifically for sex estimation on the pelvis (e.g., Klales et al., 2012). However, the current research within forensic anthropology and cognitive bias in sex estimation also shows (similar to other research within forensic science) that, even though the methods used in forensic anthropology are considered "foundationally valid" and in principle reliable, there are still arguably some unconscious factors that could affect the interpretation process. This is not unique to the discipline of forensic anthropology and sex estimation alone, but rather a phenomenon shown within any discipline where there is an element of human interpretation and subjective analysis involved (Risinger, Saks, Thompson, & Rosenthal, 2002).

Future directions

The context-sensitive nature of bioarchaeological and forensic cases means that human interpretations are highly important. Humans are still critical for interpreting the results of sensitive and accurate analytical techniques, as well as for classifying and identifying evidence within the forensic science process. This creates complexities and controversy regarding how to best deal with human factors that could cause interpretation issues (Morgan, 2017a).

Forensic anthropologists work in a variety of professional contexts, and tackling potential cognitive and contextual affects within the discipline is not an easy task. This is due to the complexities of the decision-making involved, which must often be made in line with existing policies or procedures. Many of the current proposed solutions are targeted at high-volume laboratories with much of the debate focusing on managing and blinding experts to task irrelevant information (e.g., extraneous information from a suspect's criminal record, eyewitness identifications, confessions and other lines of evidence) that could potentially cause bias (Dror, 2018).

Arguably, forensic anthropologists work within a "high-context" environment. For example, forensic anthropologists might be on site helping to preserve, excavate, and document the skeleton in situ (Dirkmaat et al., 2008). This is of importance as the expertise and knowledge of forensic anthropologists on site can significantly aid in the outcome of a death investigation and mitigate the potential for the loss of important information pertinent to the anthropological assessment of the remains (Dirkmaat et al., 2008). In addition, a large and still growing number of forensic anthropologists work within medical examiners' offices with and (in some cases) under the direction of pathologists. The communications between forensic anthropologists and the pathologist are, indeed, important, especially when trying to establish a thorough report on manners/causes of death with regard to trauma analysis.

However, these conditions could arguably also create an environment in which forensic anthropologists are exposed to significant number of contexts that, in some cases, might affect the interpretation process (Hedenstierna-Jonson et al., 2017; Nakhaeizadeh et al., 2018). Although it is important to utilize a combination of different types of evidence in the creation of a biological profile, this carries the risk of the anthropologist being, in some cases, exposed to "extraneous information." Furthermore, the data that forensic anthropologists might work with can also, in some cases, be ambiguous in nature especially when working with fragmented and poorly preserved skeletons. As research has shown that cognitive biases tend to affect our interpretations, perceptions, and memory, especially when dealing with ambiguous and difficult decisions (Kassin et al., 2013), these types of scenarios could arguably put the forensic anthropologist in a greater risk of "cognitive contamination."

In contrast to some forensic disciplines where a single repeated analysis is commonly conducted, forensic anthropologists perform multiple tests in sequence to reach a conclusion (Warren et al., 2018; Winburn, 2018). For example, in establishing biological sex, the forensic anthropologist may first look at the pelvis and the skull followed by a further examination on the long bones, applying multiple metric and morphological analyses. A potential problem could arise if the initial analyses (or traits) subconsciously affect how the next feature is interpreted when conducting multiple tests in sequence. Regardless of whether morphological or metric assessments are performed first, there is a risk of each type of test informing subsequent tests (Dror et al., 2017). In other words, there is a risk of always having a priori knowledge when conducting a subsequent assessment. Many of the assessments within forensic anthropology are sex-specific, meaning that if there are cognitive interpretation issues arising during the stage of sex assessment, arguably the interpretation of age at death could be exposed to biased evaluations as well.

Therefore, finding an appropriate balance between the risk and benefits of enacting solutions that seek to deal with the issues of cognitive and contextual biases is not an easy undertaking. These variables affecting our interpretations are not always possible to control and can in a forensic context prove to be problematic to measure and decipher. In forensic anthropology, deciding what is influential and biasing could arguably depend on the nature of the task, the ambiguity level of given characteristics being interpreted, the difficulty of the judgment, and the strength of the context in which the decision is made. Therefore, furthering our understanding of the cognitive processes in place in decision-making within forensic anthropology (and the wider forensic science disciplines) by undertaking more empirical research will generate data that can aid improving understanding of which factors lead to and influence the decision-making process.

There are two critical areas to consider when seeking to address the issues that arise from the effects of context and bias on decision-making: how we present decisions, and the role of education programs. First, while reducing the opportunities for exposure of decision-makers to extraneous information and potentially biasing contexts is valuable, it also needs to be recognized that decision-making is never free from being influenced by intrinsic and extrinsic factors. Therefore, there is great value in developing frameworks for presenting decisions that are reached in forensic reconstructions. Frameworks that allow transparency in presenting the key factors that contributed to a decision, and how that decision was reached, are a good first step (Morgan, 2017a). If these approaches can document prior beliefs, knowledge, and experience of the decision-maker as well as incorporate the critical extrinsic factors specific to the case and/or decision in a holistic manner, there is potential to increase the clarity of how a decision has been reached.

Second, change needs to stem from the bottom-up, where decision-making theories are incorporated in research and practice-led teaching at an early stage. Developing decision architecture with a greater understanding of how decisions are made will foster a culture of change (Dror, 2016; Morgan, 2017b). Embedding the inclusion of decision-making as part of the forensic science process through education and training is currently lacking within the educational system in forensic anthropology. Indeed, there is still a distinct lack of clarity just how the body of knowledge concerning the application of decision theories within forensic science can be beneficial in the educational process. Recognizing the role that cognition plays in the collection, analysis, interpretation, and presentation of evidence and the role of expertise within that process will enable the forensic anthropological community to address the concerns with regard to the issue of interpretations raised by the National Academy of Sciences report (2009) and PCAST report (2016), among others.

Conclusion

The forensic anthropological domain has come far in the development of the discipline. However, a better understanding of the underlying processes of the decisions being made and the extent to which contextual influences occur in forensic anthropology broadly, and in sex estimations specifically, need to be acknowledged and addressed. Future studies within forensic anthropology (as well as other forensic science disciplines) may need to explore the broader aspect of decision theories to fully comprehend not only how examiners reach conclusions but also how research in cognition could enhance forensic anthropological practice and procedures. Engaging with cognitive research will allow forensic practitioners to develop frameworks for communicating how an inference was reached and the basis for the conclusions made, while incorporating the potential for cognitive influences that may have had an impact on the inference and/or conclusion (Morgan, 2017b) This will allow forensic scientists to better understand and document the decision-making procedures and the role of expertise along with their limitations when making and communicating forensic interpretations (Edmond et al., 2017).

Equally, when bias detrimentally impacts a conclusion, it is often not limited to only one line of evidence that is affected. Bias snowball effect often causes different lines of evidence to bias each other (Dror, 2018). Hence, it is so important to minimize bias impacting forensic anthropology. It is only by acknowledging the limitations and uncertainties inherent in subjective decision-making that the forensic anthropology community can begin to develop a more transparent culture, embracing a dialogue that openly explores decision-making within the forensic process, determining where issues exist, increasing understanding of the human interpretation processes involved, and finding ways in which decision-making processes can be improved.

References

Aarts, H., & Dijksterhuis, A. (2002). Category activation effects in judgment and behavior: The moderating role of perceived comparability. *British Journal of Social Psychology*, *41*, 123–138.

Anderson, J. R. (2000). *Cognitive psychology and its implications* (5th ed.). United State of America: Worth Publishers and WH Freeman.

Archer, M. S., & Wallman, J. F. (2016). Context effects in forensic entomology and use of sequential unmasking in casework. *Journal of Forensic Sciences*, *61*(5), 1270–1277.

Ask, K., & Granhag, P. A. (2005). Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. *Journal of Investigative Psychology and Offender Profiling*, *2*(1), 43–63.

Balcetis, E., & Dunning, D. (2006). See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology*, *91*(4), 612–625.

Bandura, A. (1999). Social cognitive theory: A agentic perspective. *Asian Journal of Social Psychology*, *2*, 21–41.

Bhalla, M., & Proffitt, D. R. (1999). Visual-motor recalibration in geographical slant perception. *Journal of Experimental Psychology. Human Perception and Performance*, *25*(4), 1076–1096.

Bieber, P. (2012). Measuring the impact of cognitive bias in fire investigation. In *International symposium on fire investigation. Science and technology* (pp. 3–15).

Biedermann, A., Bozza, S., & Taroni, F. (2016). The decisionalization of individualization. *Forensic Science International*, *266*, 29–38.

Biwasaka, H., Aoki, Y., Tanijiri, T., Sato, K., Fujita, S., Yoshioka, K., et al. (2009). Analyses of sexual dimorphism of contemporary Japanese using reconstructed three-dimensional CT images–curvature of the best-fit circle of the greater sciatic notch. *Legal Medicine*, *11*(1), 260–262.

Bressan, P., & Dal Martello, M. F. (2002). Talis pater, Talis filius: Perceived resemblance and the belief in genetic relatedness. *Psychological Science*, *13*, 213–218.

Byers, S. N. (2010). *Introduction to forensic anthropology*. Taylor & Francis.

Cattaneo, C., Di Martino, S., Scali, S., Craig, O. E., Grandi, M., & Sokol, R. J. (1999). Determining the human origin of fragments of burnt bone: A comparative study of histological, immunological and DNA techniques. *Forensic Science International*, *102*(2–3), 181–191.

Christensen, A. M. (2004). The impact of Daubert: Implications for testimony and research in forensic anthropology (and the use of frontal sinuses in personal identification). *Journal of Forensic Sciences*, *49*(3), 427–430.

Christensen, A. M., & Crowder, C. M. (2009). Evidentiary standards for forensic anthropology. *Journal of Forensic Sciences*, *54*(6), 1211–1216.

Clark, M. A., Guatelli-Steinberg, D., Hubbe, M., & Stout, S. (2016). Quantification of maxillary dental arcade curvature and the estimation of biological ancestry in forensic anthropology. *Journal of Forensic Sciences*, *61*(1), 141–146.

Dawes, R. M. (1994). *House of cards psychology and psychotherapy built on myth*. New York: Free Press.

Dirkmaat, D. C., & Cabo, L. L. (2012). Forensic anthropology: Embracing the new paradigm. In D. C. Dirkmaat (Ed.), *A companion to forensic anthropology* (pp. 3–40): Blackwell Publishing Ltd.

Dirkmaat, D. C., Cabo, L. L., Ousley, S. D., & Symes, S. A. (2008). New perspectives in forensic anthropology. *American Journal of Physical Anthropology*, *137*(47), 33–52.

Dror, I. E. (2011). The paradox of human expertise: Why experts get it wrong. In N. Kapur (Ed.), *The paradoxical brain* (pp. 177–188). Cambridge: Cambridge University Press.

Dror, I. E. (2016). A hierarchy of expert performance (HEP). *Journal of Applied Research in Memory and Cognition*, *5*, 121–127.

Dror, I. E. (2018). Biases in forensic experts. *Science*, *360*(6386), 243.

Dror, I. E., Champod, C., Langenburg, G., Charlton, D., Hunt, H., & Rosenthal, R. (2011). Cognitive issues in fingerprint analysis: Inter-and intra-expert consistency and the effect of a "target" comparison. *Forensic Science International*, *208*(1–3), 10–17.

Dror, I. E., & Charlton, D. (2006). Why experts make errors. *Journal of Forensic Identification*, *56*(4), 600–616.

Dror, I. E., Charlton, D., & Péron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, *156*(1), 74–78.

Dror, I. E., & Hampikian, G. (2011). Subjectivity and bias in forensic DNA mixture interpretation. *Science and Justice*, *51*(4), 204–208.

Dror, I. E., Morgan, R. M., Rando, C., & Nakhaeizadeh, S. (2017). The bias snowball and the bias cascade effects: Two distinct biases that may impact forensic decision making. *Journal of Forensic Sciences*, *62*(3), 832–833.

Dror, I. E., Peron, A. E., Hind, S. L., & Charlton, D. (2005). When emotions get the better of us: The effect of contextual top-down processing on matching fingerprints. *Applied Cognitive Psychology*, *19*(6), 799–809.

Đurić, M., Rakočević, Z., & Đonić, D. (2005). The reliability of sex determination of skeletons from forensic context in the Balkans. *Forensic Science International*, *147*(2–3), 159–164.

Earwaker, H., Morgan, R. M., Harris, A. J., & Hall, L. J. (2015). Fingermark submission decision-making within a UK fingerprint laboratory: Do experts get the marks that they need? *Science & Justice*, *55*(4), 239–247.

Edmond, G., Towler, A., Growns, B., Ribeiro, G., Found, B., White, D., et al. (2017). Thinking forensics: Cognitive science for forensic practitioners. *Science & Justice*, *57*(2), 144–154.

Effros, B. (2000). Skeletal sex and gender in Merovingian mortuary archaeology. *Antiquity*, *74*(285), 632–639.

Elstein, A. S. (1999). Heuristics and biases: Selected errors in clinical reasoning. *Academic Medicine Journal of the Association of American Medical Colleges*, *74*(7), 791–794.

Evans, J. S. B. T., & Pollard, P. (1990). Belief bias and problem complexity in deductive reasoning. *Advances in Psychology*, *68*, 31–154.

Found, B. (2015). Deciphering the human condition: The rise of cognitive forensics. *Australian Journal of Forensic Sciences*, *47*(4), 386–401.

Found, B., & Ganas, J. (2013). The management of domain irrelevant context information in forensic handwriting examination casework. *Science & Justice*, *53*(2), 154–158.

Garvin, H. M. (2012). Adult sex determination. In D. C. Dirkmaat (Ed.), *A companion to forensic anthropology*: Blackwell Publishing Ltd.

Gianelli, P. C. (2007). Confirmation bias. *Criminal Justice*, *22*(3), 60–61.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482.

Girotto, V., & Politzer, G. (1990). Conversational and world knowledge constraints on deductive reasoning. *Advances in Psychology*, *68*, 87–107.

Government Chief Scientific Advisor. (2015). *Forensic science and beyond: Authenticity, provenance and assurance, evidence and case studies*. https://www.gov.uk/government/publications/forensic-science-and-beyond.

Grivas, C. R., & Komar, D. A. (2008). Kumho, Daubert, and the nature of scientific inquiry: Implications for forensic anthropology. *Journal of Forensic Sciences*, *53*(4), 771–776.

Guyomarc'h, P., & Bruzek, J. (2011). Accuracy and reliability in sex determination from skulls: A comparison of Fordisc® 3.0 and the discriminant function analysis. *Forensic Science International*, *208*(1–3), 180–e1.

Hamilton, D. L., & Zanna, M. P. (1974). Context effects in impression formation: Changes in connotative meaning. *Journal of Personality and Social Psychology*, *29*(5), 649–654.

Hedenstierna-Jonson, C., Kjellstrom, A., Zachrisson, T., Krzewinska, M., Sobrado, V., Price, N., et al. (2017). A female Viking warrior confirmed by genomics. *American Journal of Physical Anthropology*, *164*, 853–860.

Hefner, J. T. (2009). Cranial nonmetric variation and estimating ancestry. *Journal of Forensic Sciences*, *54*(5), 985–995.

Hefner, J. T., & Ousley, S. D. (2014). Statistical classification methods for estimating ancestry using morphoscopic traits. *Journal of Forensic Sciences*, *59*(4), 883–890.

Hoppitt, L., Mathews, A., Yiend, J., & Mackintosh, B. (2010). Cognitive mechanisms underlying the emotional effects of bias modification. *Applied Cognitive Psychology*, *24*(3), 312–325.

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, *2*(1), 42–52.

Kemkes-Grottenthaler, A., Löbig, F., & Stock, F. (2002). Mandibular ramus flexure and gonial eversion as morphologic indicators of sex. *HOMO – Journal of Comparative Human Biology*, *53*(2), 97–111.

Kerstholt, J., Eikelboom, A., Dijkman, T., Stoel, R., Hermsen, R., & van Leuven, B. (2010). Does suggestive information cause a confirmation bias in bullet comparisons? *Forensic Science International*, *198*(1–3), 138–142.

Khaneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Khaneman (Eds.), *Heruistic and biases the psychology of intuitive judgment* (pp. 49–81): Cambridge University Press.

Kimmerle, E. H., Ross, A., & Slice, D. (2008). Sexual dimorphism in America: Geometric morphometric analysis of the craniofacial region. *Journal of Forensic Sciences*, *53*(1), 54–57.

Klales, A. R. (2013). Current practices in physical anthropology for sex estimation in unidentified, adult individuals. In *Proceedings of the 82nd meeting of the American Association for Physical Anthropology, January* pp. 9–13.

Klales, A. R., & Lesciotto, K. M. (2016). The "science of science": Examining bias in forensic anthropology. In *Proceedings of the 68th annual scientific meeting of the American Academy of forensic sciences, February*.

Klales, A. R., Ousley, S. D., & Vollner, J. M. (2012). A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *American Journal of Physical Anthropology*, *149*(1), 104–114.

Krishan, K., Chatterjee, P. M., Kanchan, T., Kaur, S., Baryah, N., & Singh, R. K. (2016). A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. *Forensic Science International*, *261*, 165–e1.

Kukucka, J., & Kassin, S. M. (2014). Do confessions taint perceptions of handwriting evidence? An empirical test of the forensic confirmation bias. *Law and Human Behaviour*, *38*(3), 256.

Kukucka, J., Kassin, S. M., Zapf, P. A., & Dror, I. E. (2017). Cognitive bias and blindness: A global survey of forensic science examiners. *Journal of Applied Research in Memory and Cognition*, *6*(4), 452–459.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480.

Laber, T., Kish, P., Taylor, M., Owens, G., Osborne, N., & Curran, J. (2014). *Reliability assessment of current methods in bloodstain pattern analysis*. National Institute of Justice, US Department of Justice.

Langenburg, G., Champod, C., & Wertheim, P. (2009). Testing for potential contextual bias effects during the verification stage of the ACE-V methodology when conducting fingerprint comparisons. *Journal of Forensic Sciences*, *54*(3), 571–582.

Langley, N. R., Dudzik, B., & Cloutier, A. (2018). A decision tree for nonmetric sex assessment from the skull. *Journal of Forensic Sciences*, *63*(1), 31–37.

Lottering, N., MacGregor, D. M., Meredith, M., Alston, C. L., & Gregory, L. S. (2013). Evaluation of the Suchey–Brooks method of age estimation in an Australian subpopulation using computed tomography of the pubic symphyseal surface. *American Journal of Physical Anthropology*, *150*(3), 386–399.

Mahakkanukrauh, P., Ruengdit, S., Tun, S. M., Case, D. T., & Sinthubua, A. (2017). Osteometric sex estimation from the os coxa in a Thai population. *Forensic Science International*, *271*, 127–e1.

Mahfouz, M., Badawi, A., Merkl, B., Fatah, E. E. A., Pritchard, E., Kesler, K., … Jantz, L. (2007). Patella sex determination by 3D statistical shape models and nonlinear classifiers. *Forensic Science International*, *173*(2–3), 161–170.

Mattijssen, E. J. A. T., Kerkhoff, W., Berger, C. E. H., Dror, I. E., & Stoel, R. D. (2016). Implementing context information management in forensic casework: Minimizing contextual bias in firearms examination. *Science & Justice*, *56*(2), 113–122.

Miller, L. S. (1984). Bias among forensic document examiners: A need for procedural change. *Journal of Police Science & Administration*, *12*(4), 407–411.

Miller, L. S. (1987). Procedural bias in forensic science examinations of human hair. *Law and Human Behavior*, *11*(2), 157–163.

Morgan, R. M. (2017a). Conceptualising forensic science and forensic reconstruction. Part I: A conceptual model. *Science & Justice*, *57*(6), 455–459.

Morgan, R. M. (2017b). Conceptualising forensic science and forensic reconstruction. Part II: The critical interaction between research, policy/law and practice. *Science & Justice*, *57*(6), 460–467.

Morgan, R. M., Earwaker, H., Nakhaeizadeh, S., Dror, I., Harris, A., & Rando, C. (2019). Interpretation of forensic evidence: decision-making under uncertainty (at every step of the forensic science process). In R. Wortley, A. Sidebottom, G. Laycock, & N. Tilley (Eds.), *Handbook of crime science*. Abingdon, Oxon: Routledge.

Naikmasur, V. G., Shrivastava, R., & Mutalik, S. (2010). Determination of sex in South Indians and immigrant Tibetans from cephalometric analysis and discriminant functions. *Forensic Science International*, *197*(1–3), 122–e1.

Nakhaeizadeh, S., Dror, I. E., & Morgan, R. M. (2014). Cognitive bias in forensic anthropology: Visual assessment of skeletal remains is susceptible to confirmation bias. *Science & Justice*, *54*(3), 208–214.

Nakhaeizadeh, S., Hanson, I., & Dozzi, N. (2014). The power of contextual effects in forensic anthropology: A study of biasability in the visual interpretations of trauma analysis on skeletal remains. *Journal of Forensic Sciences*, *59*(5), 1177–1183.

Nakhaeizadeh, S., & Morgan, R. M. (2015). Forensic anthropology and cognitive bias. In A. Jamieson & A. Moenssens (Eds.), *Wiley encyclopedia of forensic science*. Chichester, UK: John Wiley.

Nakhaeizadeh, S., Morgan, R. M., Rando, C., & Dror, I. E. (2018). Cascading bias of initial exposure to information at the crime scene to the subsequent evaluation of skeletal remains. *Journal of Forensic Sciences*, *63*(2), 403–411.

National Research Council. (2009). *Strengthening forensic science in the United States: A path forward*. National Academies Press.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175.

Ordonez, L., & Benson, L., III. (1997). Decisions under time pressure: How time constraint affects risky decision making. *Organizational Behavior and Human Decision Processes*, *71*(2), 121–140.

Osborne, N. K., Taylor, M. C., & Zajac, R. (2016). Exploring the role of contextual information in bloodstain pattern analysis: A qualitative approach. *Forensic Science International*, *260*, 1–8.

Osborne, N. K., Woods, S., Kieser, J., & Zajac, R. (2014). Does contextual information bias bitemark comparisons? *Science & Justice*, *54*(4), 267–273.

Page, M., Taylor, J., & Blenkin, M. (2012). Context effects and observer bias—Implications for forensic odontology. *Journal of Forensic Sciences*, *57*(1), 108–112.

Phenice, T. W. (1969). A newly developed visual method of sexing the os pubis. *American Journal of Physical Anthropology*, *30*(2), 297–301.

President's Committee of Advisors on Science and Technology. (2016). *Forensic science in criminal courts: ensuring scientific validity of feature comparison methods*. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

Risinger, D. M., Saks, M. J., Thompson, W. C., & Rosenthal, R. (2002). The Daubert/Kumho implications of observer effects in forensic science: Hidden problems of expectation and suggestion. *California Law Review*, *90*, 1.

Schiffer, B., & Champod, C. (2007). The potential (negative) influence of observational biases at the analysis stage of fingermark individualisation. *Forensic Science International*, *167*(2–3), 116–120.

Spradley, M. K., & Jantz, R. L. (2011). Sex estimation in forensic anthropology: Skull versus postcranial elements. *Journal of Forensic Sciences*, *56*(2), 289–296.

Stanovich, K. E. (2014). Assessing cognitive abilities: Intelligence and more. *Journal of Intelligence*, *2*(1), 8–11.

Sternberg, R. J. (Ed.), (2008). *Why smart people can be so stupid*: Yale University Press.

Stevenage, S. V., & Bennett, A. (2017). A biased opinion: Demonstration of cognitive bias on a fingerprint matching task through knowledge of DNA test results. *Forensic Science International*, *276*, 93–106.

Steyn, M., Pretorius, E., & Hutten, L. (2004). Geometric morphometric analysis of the greater sciatic notch in South Africans. *HOMO – Journal of Comparative Human Biology*, *54*(3), 197–206.

Stoel, R. D., Dror, I. E., & Miller, L. S. (2014). Bias among forensic document examiners: Still a need for procedural changes. *Australian Journal of Forensic Sciences*, *46*(1), 91–97.

Thomas, R. M., Parks, C. L., & Richard, A. H. (2016). Accuracy rates of sex estimation by forensic anthropologists through comparison with DNA typing results in forensic casework. *Journal of Forensic Sciences*, *61*(5), 1307–1310.

Thompson, W. C., & Cole, S. A. (2007). Psychological aspects of forensic identification evidence. In *Expert psychological testimony for the courts* (pp. 31–68): Lawrence Erlbaum Associates Publishers.

van den Eeden, C. A., de Poot, C. J., & Van Koppen, P. J. (2016). Forensic expectations: Investigating a crime scene with prior information. *Science & Justice*, *56*(6), 475–481.

van den Eeden, C. A., de Poot, C. J., & van Koppen, P. J. (2018). The forensic confirmation bias: A comparison between experts and novices. *Journal of Forensic Sciences*.

Walker, P. (1995). Problems of preservation and sexism in sexing: Some lessons from historical collections for palaeodemographers. In S. Saunders & A. Herring (Eds.), *Grave reflections: Portraying the past through cemetery studies* (pp. 31–47). Toronto: Canadian Scholars Press.

Walker, P. L. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, *136*(1), 39–50.

Walrath, D. E., Turner, P., & Bruzek, J. (2004). Reliability test of the visual assessment of cranial traits for sex determination. *American Journal of Physical Anthropology*, *125*(2), 132–137.

Warren, M. W., Friend, A. N., & Stock, M. K. (2018). Navigating cognitive bias in forensic anthropology. In *Forensic anthropology: Theoretical framework and scientific basis* (pp. 39–51): John Wiley & Sons Ltd.

Weiss, K. M. (1972). On the systematic bias in skeletal sexing. *American Journal of Physical Anthropology*, *37*(2), 239–249.

Winburn, A. P. (2018). Subjective with a capital S? Issues of objectivity in forensic anthropology. In *Forensic anthropology: Theoretical framework and scientific basis* (pp. 19–37): John Wiley & Sons Ltd.

Wyer, R. S., & Srull, T. K. (1986). Human cognition in its social context. *Psychological Review*, *93*(3), 322.

CHAPTER 21

Sex determination using DNA and its impact on biological anthropology

Richard M. Thomas
Federal Bureau of Investigation, Laboratory Division, Quantico, VA, United States

Introduction

Prior to the era of DNA examinations, forensic anthropologists were the primary source of information regarding the sex of human skeletal remains. Innovations in DNA extraction, amplification, and analyses have provided a molecular method to complement the anthropological analysis. Especially integral was the discovery of differences in the amelogenin gene, which encodes enamel production, between the X and Y chromosomes. Using cloning techniques, Nakahori, Takenaka, and Nakagome (1991) localized the amelogenin gene to the sex chromosomes; and with the advent of the polymerase chain reaction (Mullis & Faloona, 1989; Saiki et al., 1988), researchers developed robust and fairly simple amplification techniques for determining the presence of X and Y chromosome versions of this gene. This technique exploited the fact that insertion/deletion polymorphisms between the X and Y chromosomes lead to differently sized amplicons, which can easily be visualized by size separation using gel electrophoresis (Bailey, Affara, & Ferguson-Smith, 1992; Mannucci, Sullivan, Ivanov, & Gill, 1994; Nakahori, Hamano, Iwaya, & Nakagome, 1991; Sullivan, Mannucci, Kimpton, & Gill, 1993). Therefore, the presence of a single band indicates that only X chromosomes are present (female), and the presence of two bands indicates that both X and Y chromosomes are present (male). Size discrepancy based on a 6 bp insertion/deletion site of the amelogenin gene was the primary basis for sex determination in the first commercially available DNA profiling kits, which were widely used in forensic laboratories (Sullivan et al., 1993). However, more recent profiling kits also include assays for a separate insertion/deletion polymorphism and a short tandem repeat (STR) locus found only on the Y chromosome (Ludeman et al., 2018). Additionally, analysis of other STRs (Nakahori, Mitani, Yamada, & Nakagome, 1986; Palmirotta et al., 1997; Witt & Erickson, 1989), other insertions and deletions (Daskalaki, Anderung, Humphrey, & Götherström, 2011), or single nucleotide polymorphisms (SNPs; Gibbon, Paximadis, Štrkalj, Ruff, & Penny, 2009) on the X or Y chromosome can also be used to determine sex. However, in assays of DNA from the Y chromosome, the lack of amplified material cannot always distinguish between a female individual and a failure of the DNA analysis (Santos, Pandya, & Tyler-Smith, 1998). DNA

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00021-3

mutations or population-specific polymorphisms may also lead to incorrect sex determination in certain cases (Ou et al., 2012).

More recently, methodological improvements using high-throughput shotgun sequencing, which determine sex based upon the amount of sequence reads to the X and Y chromosomes, have been developed (Mittnik, Wang, Svoboda, & Krause, 2016; Skoglund, Storå, Götherström, & Jakobsson, 2013). These techniques are useful for poor-quality samples where allelic dropout and contamination are of particular concern, including in ancient DNA studies such as the sex determination of an Egyptian mummy head from approximately 2000 BCE (Loreille et al., 2018).

It should be noted that despite the robustness of DNA-based sex determination assays, there are various issues that may complicate an overly simplified male vs. female dichotomy. It should be mentioned that the chromosomal makeup of an individual does not universally translate to an individual's gender, which is important since it is the reported gender that is often used for comparison in missing person cases (see Chapter 4 in this volume). Issues such as androgen insensitivity, gonadal dysgenesis, transgender individuals, and trisomy of sex chromosomes may lead to a reported gender, which does not align with the simplistic determination of the presence of X and Y chromosomes in an individual. The statistical prevalence of these issues is high enough for them to not be simply ignored by researchers and forensic scientists (von Wurmb-Schwark, Bosinski, & Ritz-Timme, 2007).

DNA extraction from skeletal material

The development and refinement of DNA extraction techniques from skeletal material proceeded concurrently with the development of sex determination from the amelogenin gene (Hagelberg & Clegg, 1991; Hagelberg, Sykes, & Hedges, 1989). Once DNA is extracted from skeletal tissue, downstream amplification and analysis proceeds in the same manner as DNA extracted from other tissues.

Achieving success in DNA extraction from skeletal material is highly dependent on choosing the right skeletal element to sample. Due to their inherent relative resistance to taphonomic processes and resulting ability to protect internal DNA, teeth are often favored above other skeletal tissues, if they are available. The teeth may be easily removed in some cases but—in other cases, especially for the tenacious roots of upper molars—may require cutting of the alveolar region for extraction.

Long bones, such as the femur and tibia, are also favored because of their large amount of available compact bone, especially in cases with taphonomic damage. However, studies have shown that a wide variety of skeletal elements provide enough DNA for analysis (Edson et al., 2009; Edson, Ross, Coble, Parson, & Barritt, 2004), and in cases where taphonomic weathering is minimal, such as mass disasters or burials, smaller intact bones with more trabecular bone may provide a larger DNA yield and be favorable because of the ease of sampling (Mundorff, Davoren, & Weitz, 2013) As a result, if at least a partial skeleton is present, the choice of bone for extraction is typically based on a variety of factors, including preservation, amount of compact bone available, presence of trauma or pathology, and ease of sampling.

Once the bone is chosen for extraction, the first step is to properly document the bone prior to sampling. This is usually done through photography, although more advanced documentation such as CT scanning may be useful. After documentation, the bone is cleaned to remove contaminant DNA or inhibitors (such as soil) from the surface. This is usually done through scrubbing with clean nylon brushes or paper towels, cleaning with diluted bleach, or sanding of the bone surface (Kemp & Smith, 2005). The nature of surface contaminants, the amount of compact bone available, and the surface topography of the bone itself dictate which cleaning procedure will be most effective.

Although, in some cases, the entire bone or bone fragment may be consumed, most often the bone or tooth is cut to provide the material for DNA extraction. This usually involves the use of rotary tools with attached circular blades. Four cuts into the bone produces a square or rectangular "window" of bone for extraction, although a wedge of bone can be removed using two angular cuts (from the anterior crest of the tibia, for example). Other times, just a portion of a smaller bone or rib section is used. In all cases, sampling should avoid any areas of perimortem trauma to the skeleton, or any areas containing information that may be useful for identification, such as age, sex, or ancestry indicators, healed trauma, pathological conditions, or skeletal anomalies. The Scientific Working Group for Forensic Anthropology (SWGANTH) has produced guidelines to anthropologists for the preparation and sampling of skeletal material for DNA analysis, although consultation with the DNA laboratory is a key part of this process.

Once the sampling process is complete, the skeletal material is reduced to powder for the extraction of DNA. This can be done using cryogenic mills or grinders, where the bone or tooth is placed in a chamber with metal endcaps and a metal pulverizing rod. The chamber is then placed in a liquid nitrogen bath, which freezes the bone, and the mill uses magnetic pulses to cause the rod move back and forth to pulverize the sample. The powdering process can also be done through the use of blenders (Edson et al., 2004). The collection of powder through drilling directly into the bone sometimes is done in place of cutting and powdering a bone sample (Rohland & Hofreiter, 2007).

Most commonly, the powder is then incubated in a demineralization buffer (Loreille, Diegoli, Irwin, Coble, & Parsons, 2007), which helps release cells from the crystalline structure of the bone (Götherström, Collins, Angerbjörn, & Lidén, 2002). Then the DNA is released from the cells using detergent, the spent powder is separated though centrifugation, and the DNA is isolated and cleaned using the same various procedures as DNA extracted from other tissues (Rohland & Hofreiter, 2007).

Impacts of DNA-based sex determination on the field of biological anthropology

The ability to determine sex based on DNA from skeletal remains has had a major impact on the fields of archaeology and biological anthropology. This is especially true for individuals where sex estimation based on morphological or metric analysis was not possible, such as juvenile remains and partial remains. An early study (Faerman et al., 1995) focusing on a wide variety of archaeological remains from varying time periods and geographic areas showed the promise of the new DNA sex determination techniques for biological anthropology. Individual studies have also compared the morphologically estimated or known sex to DNA sex determination in areas such as Sweden (Götherström, Liden, Ahlström, Källersjö, & Brown, 1997), northern Russia (Ovchinnikov, Ovtchinnikova, Druzina, Buzhilova, & Makarov, 1998), Turkey (Matheson & Loy, 2001), and Germany (Hummel, Bramanti, Finke, & Herrmann, 2000).

In the following years, several individual studies have exploited the new DNA techniques to determine the sex of individual skeletons and, as a result, attempted to answer questions about specific anthropological topics such as marriage patterns, burial patterns, and differential patterns of mortality rates, disease, diet, status, and material possessions (Kaestle & Horsburgh, 2002). Studies included skeletal remains from archaeological sites around the world, including: juvenile skeletons from the Netherlands (Colson, Richards, Bailey, Sykes, & Hedges, 1997); stillborn infants from Switzerland (Lassen, Hummel, & Herrmann, 2000); infanticide in Roman-era Britain (Mays & Faerman, 2001); Israel (Faerman et al., 1998); Aztec-era Mexico (De La Cruz et al., 2008); an Upper Paleolithic triple burial (Mittnik et al., 2016); children in a convent burial in Portugal (Cunha et al., 2000); a Viking-age grave in Sweden (Olausson and Gotherstrom (1998); and ancient Native American skeletons from Illinois (Stone, Milner, Pääbo, & Stoneking, 1996).

Impacts of DNA-based sex determination on forensic anthropological casework

Sex determination through DNA was quickly applied to the field of forensics (Akane et al., 1992), from the development of DNA extraction from bone to the analysis of unidentified human skeletal remains. As a result, it became possible to determine sex from bones and teeth both to confirm anthropological estimations of sex and to assist in cases where sex estimation was not possible due to a lack of appropriate skeletal material or to ambiguous results. Furthermore, the use of DNA for sex determination from skeletal material is a major example of how the field of forensic anthropology has been shifted away from merely providing biological profile information for identification purposes (Dirkmaat, Cabo, Ousley, & Symes, 2008).

An early evaluation of the accuracy of anthropological sex estimation using DNA markers showed an 88% success rate (Hummel et al., 2000). A larger, more recent review of a wide variety of forensic anthropology cases with subsequent DNA-based sex determinations has suggested that anthropological estimation of sex has an overall success rate of approximately 94.7% (Thomas, Parks, & Richard, 2016). This success rate is highly dependent on which bones were available for analysis by the anthropologist. In cases where the anthropological estimation of sex was incorrect, subsequent sex determination through the DNA provides an opportunity to revisit and update the anthropological analysis.

Since estimations of the aspects of the biological profile of skeletal remains (age, sex, ancestry, and stature) are highly dependent upon one another, having a corrected sex from DNA examinations (or definitive sex if the initial estimation was undetermined) allows the re-estimation of ancestry, age, and stature. For example, an anthropologist may choose to revisit an ancestry estimation using FORDISC (Ousley & Jantz, 2012) and select only female or male groups for comparison or provide an updated stature estimation based on the corrected sex of the remains (e.g., estimation of stature based upon a 20th-century white female database rather than a white male database). Similarly, age at death intervals based on morphological characteristics of various portions of the skeleton can be redone using the correct reference database (i.e., male vs. female).

As a result, it may be beneficial for an anthropologist to revisit their cases after DNA examinations have been completed to see if their sex estimation was correct. Likewise, in rare cases where a facial approximation is completed prior to DNA analysis, it may be necessary to reproduce the facial approximation as a male instead of a female, for example.

Conclusion

It is clear that DNA-based sex determination of skeletal remains has revolutionized the ability of anthropologists to confirm morphological and metric sex estimations and to determine sex for remains where traditional techniques are not useful, such as juvenile remains, fragmentary remains, or remains with ambiguous sex characteristics. This has allowed new avenues of exploration for past population studies and for certain forensic cases. It should be mentioned that DNA cannot, however, be extracted from all remains, such as those with extreme taphonomic degradation or generally poor DNA preservation. In these cases, traditional anthropological techniques will continue to be the sole means of assessing the sex of the individual.

Legal disclaimer

This publication number 20-57 of the FBI Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for information only and inclusion does not imply endorsement by the FBI or the U.S. Government. The views expressed are those of the authors and do not necessarily reflect the official policy or position of the FBI or the U.S. Government.

References

Akane, A., Seki, S., Shiono, H., Nakamura, H., Hasegawa, M., Kagawa, M., et al. (1992). Sex determination of forensic samples by dual PCR amplification of an XY homologous gene. *Forensic Science International*, *52*(2), 143–148.

Bailey, D. M., Affara, N. A., & Ferguson-Smith, M. A. (1992). The XY homologous gene amelogenin maps to the short arms of both the X and Y chromosomes and is highly conserved in primates. *Genomics*, *14*(1), 203–205.

Colson, I. B., Richards, M. B., Bailey, J. F., Sykes, B. C., & Hedges, R. E. M. (1997). DNA analysis of seven human skeletons excavated from the Terp of Wijnaldum. *Journal of Archaeological Science*, *24*(10), 911–917.

Cunha, E., Fily, M. L., Clisson, I., Santos, A. L., Silva, A. M., Umbelino, C., et al. (2000). Children at the convent: Comparing historical data, morphology and DNA extracted from ancient tissues for sex diagnosis at Santa Clara-a-Velha (Coimbra, Portugal). *Journal of Archaeological Science*, *27*(10), 949–952.

Daskalaki, E., Anderung, C., Humphrey, L., & Götherström, A. (2011). Further developments in molecular sex assignment: A blind test of 18th and 19th century human skeletons. *Journal of Archaeological Science*, *38*(6), 1326–1330.

De La Cruz, I., González-Oliver, A., Kemp, B. M., Román, J. A., Smith, D. G., & Torre-Blanco, A. (2008). Sex identification of children sacrificed to the ancient Aztec rain gods in Tlatelolco. *Current Anthropology*, *49*(3), 519–526.

Dirkmaat, D. C., Cabo, L. L., Ousley, S. D., & Symes, S. A. (2008). New perspectives in forensic anthropology. *American Journal of Physical Anthropology*, *137*(S47), 33–52.

Edson, S. M., Christensen, A. F., Barritt, S. M., Meehan, A., Leney, M. D., & Finelli, L. N. (2009). Sampling of the cranium for mitochondrial DNA analysis of human skeletal remains. *Forensic Science International: Genetics Supplement Series*, *2*(1), 269–270.

Edson, S. M., Ross, J. P., Coble, M. D., Parson, T. J., & Barritt, S. M. (2004). Naming the dead-confronting the realities of the rapid identification of degraded skeletal remains. *Forensic Science Review*, *16*(1), 63–88.

Faerman, M., Bar-Gal, G. K., Filon, D., Greenblatt, C. L., Stager, L., Oppenheim, A., et al. (1998). Determining the sex of infanticide victims from the late Roman era through ancient DNA analysis. *Journal of Archaeological Science*, *25*(9), 861–865.

Faerman, M., Filon, D., Kahila, G., Greenblatt, C. L., Smith, P., & Oppenheim, A. (1995). Sex identification of archaeological human remains based on amplification of the X and Y amelogenin alleles. *Gene*, *167*(1), 327–332.

Gibbon, V., Paximadis, M., Štrkalj, G., Ruff, P., & Penny, C. (2009). Novel methods of molecular sex identification from skeletal tissue using the amelogenin gene. *Forensic Science International: Genetics*, *3*(2), 74–79.

Götherström, A., Collins, M. J., Angerbjörn, A., & Lidén, K. (2002). Bone preservation and DNA amplification. *Archaeometry*, *44*(3), 395–404.

Götherström, A., Liden, K., Ahlström, T., Källersjö, M., & Brown, T. A. (1997). Osteology, DNA and sex identification: Morphological and molecular sex identifications of five neolithic individuals from Ajvide, Gotland. *International Journal of Osteoarchaeology*, 7(1), 71–81.

Hagelberg, E., & Clegg, J. B. (1991). Isolation and characterization of DNA from archaeological bone. *Proceedings of the Royal Society of London B*, *244*(1309), 45–50.

Hagelberg, E., Sykes, B., & Hedges, R. (1989). Ancient bone DNA amplified. *Nature*, *342*(6249), 485.

Hummel, S., Bramanti, B., Finke, T., & Herrmann, B. (2000). Evaluation of morphological sex determinations by molecular analyses. *Anthropologischer Anzeiger*, 9–13.

Kaestle, F. A., & Horsburgh, K. A. (2002). Ancient DNA in anthropology: Methods, applications, and ethics. *American Journal of Physical Anthropology*, *119*(S35), 92–130.

Kemp, B. M., & Smith, D. G. (2005). Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. *Forensic Science International*, *154*(1), 53–61.

Lassen, C., Hummel, S., & Herrmann, B. (2000). Molecular sex identification of stillborn and neonate individuals ("Traufkinder") from the burial site Aegerten. *Anthropologischer Anzeiger*, 1–8.

Loreille, O., Ratnayake, S., Bazinet, A. L., Stockwell, T. B., Sommer, D. D., Rohland, N., et al. (2018). Biological sexing of a 4000-year-old Egyptian mummy head to assess the potential of nuclear DNA recovery from the most damaged and limited forensic specimens. *Genes*, *9*(3), 135.

Loreille, O. M., Diegoli, T. M., Irwin, J. A., Coble, M. D., & Parsons, T. J. (2007). High efficiency DNA extraction from bone by total demineralization. *Forensic Science International: Genetics*, *1*(2), 191–195.

Ludeman, M. J., Zhong, C., Mulero, J. J., Lagace, R. E., Hennessy, L. K., Short, M. L., et al. (2018). Developmental validation of GlobalFiler™ PCR amplification kit: A 6-dye multiplex assay designed for amplification of casework samples. *International Journal of Legal Medicine*, *132*(6), 1555.

Mannucci, A., Sullivan, K. M., Ivanov, P. L., & Gill, P. (1994). Forensic application of a rapid and quantitative DNA sex test by amplification of the XY homologous gene amelogenin. *International Journal of Legal Medicine*, *106*(4), 190–193.

Matheson, C. D., & Loy, T. H. (2001). Genetic sex identification of 9400-year-old human skull samples from Çayönü Tepesi, Turkey. *Journal of Archaeological Science*, *28*(6), 569–575.

Mays, S., & Faerman, M. (2001). Sex identification in some putative infanticide victims from Roman Britain using ancient DNA. *Journal of Archaeological Science*, *28*(6), 555–559.

Mittnik, A., Wang, C. C., Svoboda, J., & Krause, J. (2016). A molecular approach to the sexing of the triple burial at the Upper Paleolithic Site of Dolní Věstonice. *PLoS One*, *11*(10), e0163019.

Mullis, K. B., & Faloona, F. A. (1989). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. In *Recombinant DNA methodology* (pp. 189–204). Academic Press.

Mundorff, A. Z., Davoren, J., & Weitz, S. (2013). *Developing an empirically based ranking order for bone sampling: Examining the differential DNA yield rates between human skeletal elements over increasing post mortem intervals*. Washington, DC: National Institute of Justice, Final technical report 2010-DN-BX-K229.

Nakahori, Y., Hamano, K., Iwaya, M., & Nakagome, Y. (1991). Sex identification by polymerase chain reaction using X-Y homologous primer. *American Journal of Medical Genetics*, *39*(4), 472–473.

Nakahori, Y., Mitani, K., Yamada, M., & Nakagome, Y. (1986). A human Y-chromosome specific repeated DNA family (DYZ1) consists of a tandem array of pentanucleotides. *Nucleic Acids Research*, *14*(19), 7569–7580.

Nakahori, Y., Takenaka, O., & Nakagome, Y. (1991). A human XY homologous region encodes "amelogenin" *Genomics*, *9*(2), 264–269.

Olausson, L. H., & Gotherstrom, A. (1998). Sex identification of a skeleton in a new chamber-grave from Birka. Laborativ Arkeolpgi. *Journal of Nordic Archaeological Science*, *10*, 11.

Ou, X., Chen, W., Chen, H., Zhao, F., Zheng, J., Tong, D., et al. (2012). Null alleles of the X and Y chromosomal amelogenin gene in a Chinese population. *International Journal of Legal Medicine*, *126*(4), 513–518.

Ousley, S. D., & Jantz, R. L. (2012). FORDISC 3 and statistical methods for estimating sex and ancestry. In D. C. Dirkmatt (Ed.), *A companion to forensic anthropology* (pp. 311–329). Wiley Blackwell: London, UK.

Ovchinnikov, I. V., Ovtchinnikova, O. I., Druzina, E. B., Buzhilova, A. P., & Makarov, N. A. (1998). Molecular genetic sex determination of Medieval human remains from North Russia: Comparison with archaeological and anthropological criteria. *Anthropologischer Anzeiger*, 7–15.

Palmirotta, R., Verginelli, F., Di Tota, G., Battista, P., Cama, A., Caramiello, S., et al. (1997). Use of a multiplex polymerase chain reaction assay in the sex typing of DNA extracted from archaeological bone. *International Journal of Osteoarchaeology*, 7(6), 605–609.

Rohland, N., & Hofreiter, M. (2007). Ancient DNA extraction from bones and teeth. *Nature Protocols*, *2*(7), 1756.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., et al. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, *239*(4839), 487–491.

Santos, F. R., Pandya, A., & Tyler-Smith, C. (1998). Reliability of DNA-based sex tests. *Nature Genetics*, *18*(2), 103.

Skoglund, P., Storå, J., Götherström, A., & Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. *Journal of Archaeological Science*, *40*(12), 4477–4482.

Stone, A. C., Milner, G. R., Pääbo, S., & Stoneking, M. (1996). Sex determination of ancient human skeletons using DNA. *American Journal of Physical Anthropology*, *99*(2), 231–238.

Sullivan, K. M., Mannucci, A., Kimpton, C. P., & Gill, P. (1993). A rapid and quantitative DNA sex test: Fluorescence-based PCR analysis of XY homologous gene amelogenin. *BioTechniques*, *15*(4), 636–638.

Thomas, R. M., Parks, C. L., & Richard, A. H. (2016). Accuracy rates of sex estimation by forensic anthropologists through comparison with DNA typing results in forensic casework. *Journal of Forensic Sciences*, *61*(5), 1307–1310.

von Wurmb-Schwark, N., Bosinski, H., & Ritz-Timme, S. (2007). What do the X and Y chromosomes tell us about sex and gender in forensic case analysis? *Journal of Forensic and Legal Medicine*, *14*(1), 27–30.

Witt, M., & Erickson, R. P. (1989). A rapid method for detection of Y-chromosomal DNA from dried blood specimens by the polymerase chain reaction. *Human Genetics*, *82*(3), 271–274.

CHAPTER 22

The application of medical imaging to the anthropological estimation of sex

Samantha K. Rowbotham, Soren Blau
Victorian Institute of Forensic Medicine/Department of Forensic Medicine, Monash University, Melbourne, VIC, Australia

Medical imaging is the technique and process of creating visual representations of the hard and soft tissues of the human body for the purposes of diagnosing and treating pathology and trauma. Imaging modalities can generally be classified as those that use ionizing radiation, that is, Röentgen rays (conventional X-rays) and computed tomography (CT), and those that do not, such as magnetic resonance imaging (MRI) and ultrasound. While these imaging modalities were developed for medical purposes, they have also been used to augment practice and answer research questions in the field of biological anthropology, both in archaeological and medico-legal contexts, for the last century (Franklin, Swift, & Flavel, 2016).

The ability to visually penetrate the body's soft tissues has revolutionized the anthropologist's capability to view skeletal structures in a timely and efficient manner, regardless of the antiquity and preservation of the remains (e.g., recently deceased, mummified, decomposed, burnt, embedded in concrete) and the purpose of the investigation (e.g., to address questions related to paleoanthropology, biological archaeology, mass disasters or medico-legal casework). This has greatly benefited anthropological practice and research. For example, in anthropological cases where there is remaining soft tissue, imaging negates the need to macerate the remains. Thus, as a "triaging" process, imaging is non-destructive and less labor-intensive. In addition, avoiding invasive procedures may also assist in addressing ethical concerns related to the analysis of human remains or religious objections to autopsy. The application of medical imaging to the analysis of human remains is, therefore, increasingly considered a standard component of biological anthropology, now commonly referred to as "virtual anthropology" (Christensen, Smith, Gleiber, Cunnignham, & Wescott, 2018; Davy-Jow & Decker, 2014; Dedouit et al., 2014; Franklin et al., 2016; Uldin, 2017).

Depending on the nature of the anthropological investigation, the preservation of the remains, and the infrastructure of the agencies undertaking the investigation, imaging may be used to complement traditional anthropological assessments of gross bone, or imaging may be used in lieu of an analysis of the gross bone. The use of imaging in the analysis and interpretation of human remains extends to the anthropological

Sex Estimation of the Human Skeleton
https://doi.org/10.1016/B978-0-12-815767-1.00022-5

estimation of biological sex. This chapter will examine the contributions medical imaging has made to the investigation of sexual dimorphism in anthropology; provide a review of the ways in which X-ray, CT, and MRI modalities have been implemented into anthropological estimations of sex, and consider some of the associated limitations.

Use of medical imaging in anthropological estimations of sex

The development of medical imaging modalities, and their increased application in clinical (i.e., hospital) and forensic (i.e., mortuary) settings over recent decades, has seen the inclusion of imaging as an additional tool for biological anthropologists to examine and analyze the gross bone to estimate biological sex. Traditionally, anthropological questions relating to the estimation of biological sex have been investigated using physical osteology reference collections. While these reference collections are invaluable resources for research (Henderson & Cardoso, 2018), many consist of archaeological or historical populations, with relatively few collections representative of modern individuals. This limitation is partially rectified by medical imaging datasets. Imaging data present an opportunity to develop unique reference data from contemporary individuals and, therefore, a means to more accurately address questions related to biological anthropology. Imaging data that may be available for anthropological research (pending relevant ethical approvals) is opportunistic as it comes directly from either hospitals, where scans are taken for diagnostic purposes, or from medico-legal institutions, where scans are undertaken as part of the autopsy process. Depending on the nature of the investigation, the anatomical region scanned may be either a select body area (e.g., head and neck) or the full body. These collections of images subsequently provide large digital skeletal reference collections of individuals of documented age and biological sex from modern populations. Consequently, they are invaluable resources for developing contemporary population-specific standards for sex estimation; documenting modern human variation in sexual dimorphism; and for validating pre-existing sex estimation techniques. The digital nature of these “collections” also means they have the potential to be viewed anywhere at any time (The University of New Mexico, 2018), which opens accessibility of data for research.

The advantages provided by digital collections means that there has been a gradual shift in recent research in the field of sex estimation from analyzing gross bone to examining images. A retrospective review of all papers published on sexual dimorphism in leading forensic and biological anthropology journals over the last decade shows that physical analyses of gross bone in anthropology research and practice are increasingly being replaced by the examination of medical images (Fig. 1). This trend is more rapidly developing in medico-legal contexts rather than archaeological work (see Fig. 1). This trend is undoubtedly attributable to the need for modern human reference material in medico-legal contexts that cannot be provided with historic osteology collections, and the increased use of routine postmortem imaging in many forensic medical institutes/offices of medical examiners globally.

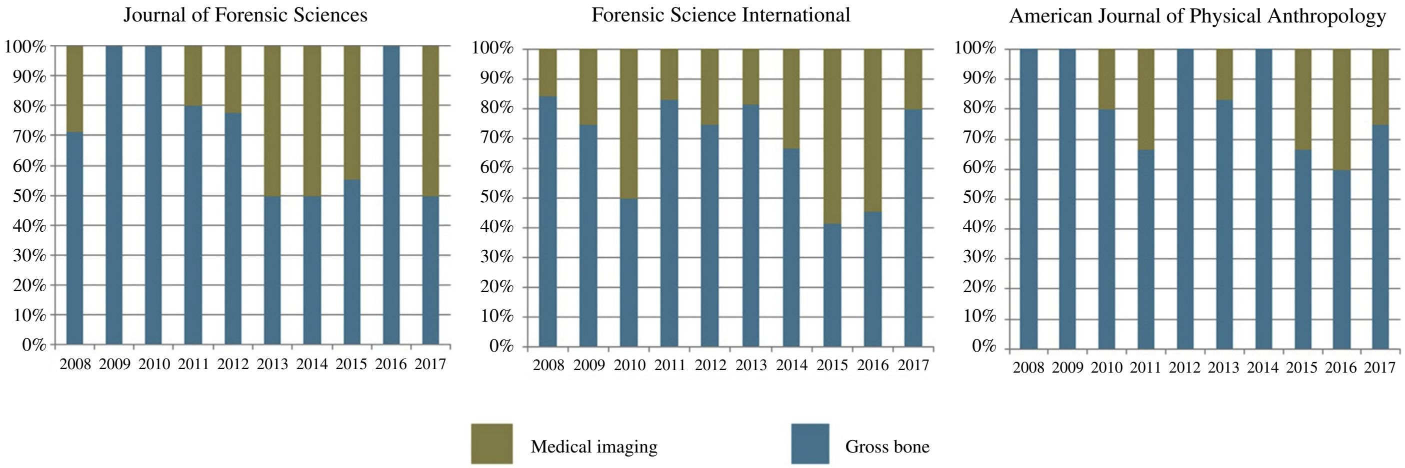


Fig. 1 Proportion of articles on sexual dimorphism in forensic and biological anthropology journals (2008–2017) that used medical imaging compared to traditional gross bone.

The premise to estimating sex from skeletal remains using medical imaging is the same as that applied when using gross bone. That is, as summarized in Rowbotham (2016), specific skeletal elements are examined that are known to demonstrate phenotypic differences between males and females. Sexually dimorphic differences are expressed through shape and size variation in skeletal anatomy and are assessed using morphological and/or metric techniques. The accuracy of such techniques varies substantially from 50% (i.e., chance) to 95% (i.e., the considered "optimal," although difficult to attain, accuracy; Krishan et al., 2016). As morphological methods are considered less objective than metric methods (although these techniques are beginning to be revised to include statistical analyses as a means to eliminate their subjectivity—Horbaly, Kenyhercz, Hubbe, and Steadman (2019); Klales, Ousley, and Vollner (2012); Walker (2008)), there has been a shift in recent years toward focusing on using metric techniques. This shift may be attributed to two reasons. First, metric methods have a strong statistical basis, which is considered more objective than morphological methods as they can be scientifically validated. Such an approach, particularly in medico-legal contexts, is more in accordance with the *Daubert* (US Supreme Court, 1993) and *President's Council of Advisors on Science and Technology* (Holdren & Lander, 2016) criteria. Second, metric techniques are potentially able to identify sexual dimorphism from skeletal elements with relatively subtle morphological dimorphic differences, as opposed to the traditionally used elements of the pelvis, skull, and long bones, which would not necessarily be visually identifiable.

X-rays

Röentgen radiation, first discovered by physicist Wilhelm Röentgen in 1895, is commonly referred to as X-radiation or conventional X-ray (Mikla & Mila, 2014). The technique discharges electromagnetic radiation through the body and, depending on the density of the tissue, a variable proportion of that radiation is absorbed by the body before passing into a radiation-sensitive film with photoreceptors (Aichinger, Dierker, Joite-Barfuß, & Säbel, 2012; Brogdon, 2011). The photoreceptors detect these differences in the densities of X-rays, often referred to as attenuations, capturing a two-dimensional (2D) superimposition of the structure of the body (Aichinger et al., 2012; Mikla & Mila, 2014). These images are typically captured in the anterior-posterior, lateral, and/or oblique planes. The hard tissues of the body primarily comprise calcium and phosphorous; these have high atomic masses and are, therefore, relatively dense tissues compared to the body's soft tissues, which have low atomic masses as they largely comprise water (i.e., hydrogen) (Mendelejeff, 1869; Mitchell, Hamilton, Steggerda, & Bean, 1945). Consequently, skeletal remains absorb a large proportion of X-ray particles. This means that osseous materials are clearly visualized on 2D X-ray images, which may be either plain film or digital. While plain-film radiographs were initially employed by archaeologists and palaeopathologists in the late 1800s to study the content of mummy bundles and later Egyptian pharaohs (Chhem & Brothwell, 2008), X-rays have also been used in more recent years to estimate biological sex from skeletal structures.

Most studies that utilize radiographs to develop sex estimation techniques are based on measurements of the skull, including dental orthopantomographs (OPGs) and panoramic radiographs. Using a variety of cephalometry measurements from radiographs and discriminant function analyses, sex estimation methods have been developed for Thai (Hsiao, Chang, & Liu, 1996), South Indian and immigrant Tibetan (Naikmasur, Shrivastava, & Mutalik, 2010), and French (Veyre-Goulet, Mercier, Robin, & Guérin, 2008) populations with reported accuracies ranging from 81.5% to 100%. Comparatively, methods that use dental radiographs to measure mandibular rami and permanent dentition to inform sexual dimorphism, showed lower accuracies. Mandibular measurements have resulted in overall low accuracies of 69% (More, Vijayvargiya, & Saha, 2017) and 75.8% (Sambhana et al., 2016), whilst measurements from dentition, although accuracies were not reported, have shown their discriminative ability was too low to be considered an acceptable misclassification error (Capitaneanu, Willems, Jacobs, Fieuws, & Thevissen, 2017). Radiographs of the head that are either anterior-posterior or lateral have also presented opportunities to identify sexually dimorphic differences in the paranasal sinuses. Skull measurements in Indian populations have identified the sinuses as sexually dimorphic in size; however, these have shown overall low accuracies of only 64.6% (Belaldavar, Kotrashetti, Hallikerimath, & Kale, 2014) and 67.59% (Sai kiran, Ramaswamy, & Khaitan, 2014) for the frontal sinuses, and 76% for the maxillary sinuses (Sidhu et al., 2014).

Chest radiographs present opportunities to investigate sex differences in the ossification patterns of costal cartilage. Sex-specific patterns of costal cartilage ossification have been identified in European populations (Rejtarová et al., 2009), and in North American populations they have been reported to be as accurate as 99% when assessed in conjunction with sternal and fourth rib dimensions (McCormick, Stewart, & Langford, 1985).

Radiographs of the bones of the upper and lower extremities have also provided a means to measure various skeletal dimensions to assess sexual dimorphism. As with the skull and chest, studies have used radiographs that were originally taken for clinical diagnoses and, therefore, comprise an eclectic mix of anatomical skeletal regions. For example, hand radiographs were used to measure dimensions of the metacarpals and proximal phalanges in a Western-Australian population; the results of which showed evidence of sexual dimorphism with an accuracy as high as 91% (DeSilva, Flavel, & Franklin, 2014). Measurements of the humerus from radiographs in a Saudi population further confirmed what has been found on gross bone—that the size of the humerus is sexually dimorphic, with an accuracy of 88.4%–94.3% (Shehri & Soliman, 2015). Similar accuracy rates for sexual dimorphism have been reported for measurements of bones of the lower extremity. Measurements of the patella and metatarsals for a contemporary Egyptian population were shown to be sexually dimorphic with an accuracy as high as 100% (Abdel Moneim, Abdel Hady, Abdel Maaboud, Fathy, & Hamed, 2008).

While some research has been undertaken on estimating sex from pre-pubescent gross human remains, as summarized in Blau and Hill (2014), the utility of these findings is questionable (Lewis, 2007). Radiographs of subadult skulls and long bones with documented known sex, however, have provided an exciting avenue for sex estimation research in this area (see Chapter 14 of this volume). Cephalometric radiographs have been used to identify sex-specific patterns of growth with juvenile craniofacial ontogeny. In North American subadults of European ancestry, Gonzalez (2012) found accuracies of 78%–89% in sexually dimorphic differences, particularly in size, of craniofacial growth from 5-year-olds until the onset of puberty. Similarly, Holton, Alsamawi, Yokley, and Froehle (2016), using cephalometric radiographs from a different longitudinal growth dataset, identified sexual dimorphism in subadult nasal shape with ontogeny. Long bone diaphyseal dimensions in South African subadults have also been measured using LODOX Statscan radiographic images (a form of low-dose digital X-ray technology) by Stull, L'Abbé, and Ousley (2017). The results from this research found sexually dimorphic differences in bone size with accuracy reported between 70% and 93% (Stull et al., 2017). It must be noted, however, that accurate sex estimations in subadults from imaging has only been attained from population-specific samples to date. O'Donnell, Berry, and Edgar (2017) applied cephalometrics to a non-specific subadult population sample and found that only 50% of the individuals were attributed to the correct sex.

Computed tomography

Computed tomography, invented by electrical engineer Sir Godfrey Hounsfield in 1972, derives from the ancient Greek work "tomos," meaning a slice or section, and thus refers to computing X-ray images of slices of the body (Mikla & Mila, 2014). A CT machine, although dependent on the specifications, may capture between 4 and 640 X-ray slices of a section of the body from a variety of angles. This is referred to as multislice CT (MSCT) or multidirectional CT (MDCT). This type of CT modality is the most commonly used form in clinical settings and is increasingly being implemented in medico-legal institutions (Oesterhelweg & Thali, 2008). Each set of multidirectional X-ray slices is stored as a DICOM (Digital Imaging and Communications in Medicine) dataset that is viewed through tailored software programs. This computer processing of slices involves electronically stitching the slices together into a series of axial, coronal, and sagittal views of the body in 2D images and "stacking" them together to create three-dimensional (3D) volume-rendered (VR) images (Fig. 2) (Brogdon, 2011).

CT follows the same physical principles as X-rays (detailed above). The different chemical compositions of structures in the human body means they each have different attenuation values. On CT, attenuation values are measured in Hounsfield units (HU). In the case of bone, the material is dense and consequently has relatively high attenuation values compared to, for example, air. These high attenuation values make bone one of the most distinctive tissues to differentiate on CT. Of the available CT modalities, MSCT is the most widely utilized in biological anthropology (Dedouit et al., 2011). This modality involves a gantry scanner, a ring that holds the radiation detector. The individual (whether alive in a clinical setting, or deceased in a mortuary context) is laid on a table that passes through the gantry. Röentgen rays are then emitted from a rotating X-ray tube in the gantry to pass through the body in the axial plane (Mikla & Mila, 2014).

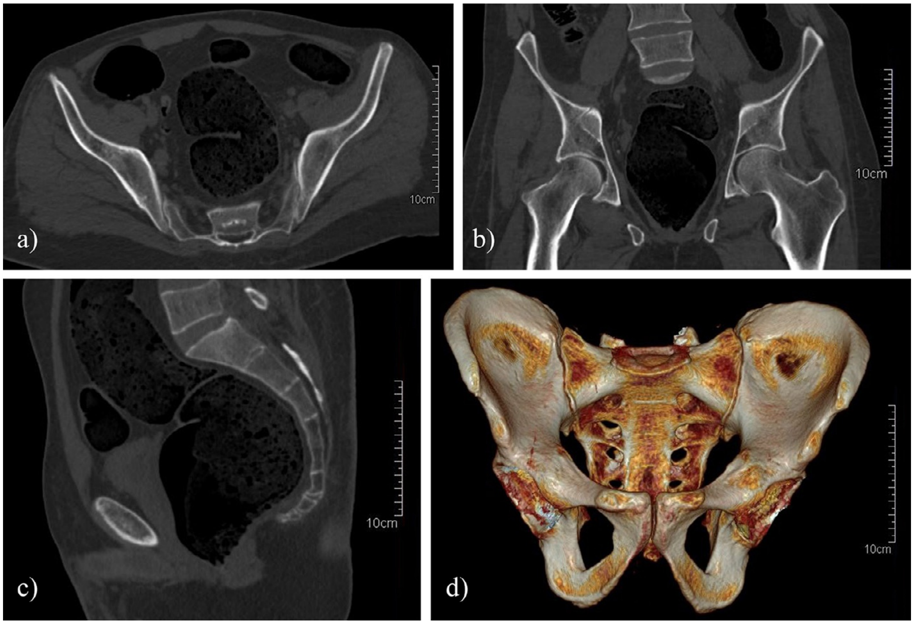


Fig. 2 Postmortem MSCT imaging of a male pelvis, viewed with Siemens Healthcare Syngo.via (VB20A_HF04) multimodality imaging software, showing the 2D reconstructed axial (A), coronal (B), and sagittal (C) slices, and the reconstructed 3D VR of the pelvic girdle, anterior aspect (D). *(Image courtesy: Victorian Institute of Forensic Medicine).*

DICOM datasets of individuals with known documented biological sex provide an invaluable resource to investigate variation in skeletal elements known to be sexually dimorphic (i.e., skull, pelvis, and long bones) using both traditional and non-traditional anthropological techniques. Using specifically tailored computer programs, studies have investigated sexual dimorphism through measuring volume and surface area of long bones (Lee, Kim, & Kwak, 2015); measuring the area of the sacrum (Zech, Hatch, Siegenthaler, Thali, & Lösch, 2012); applying curvature analyses to the innominate (Biwasaka et al., 2012); applying automated 3D global and local analyses to the crania (Abdel Fatah, Shirley, Jantz, & Mahfouz, 2014), and employing Fourier analyses to quantify the morphological characteristics of the obturator foramen (Bierry, Le Minor, & Schmittbuhl, 2010). These novel sex estimation methods further validate, and augment, current anthropological understandings of sexual dimorphism in these bones.

MSCT also offers the opportunity to investigate the usefulness of estimating sex from anatomical skeletal regions that are small and complex, and thus not easily accessible during an examination of gross bone (e.g., paranasal sinuses). Furthermore, MSCT provides the opportunity to develop novel metric techniques that are only possible with specific software programs and not with the gross bone (e.g., measuring volume). A selection of those techniques that use small anatomical skeletal regions and/or novel software analyses are documented in Table 1. These techniques are still based on the principle that there are

Table 1 Examples of novel metric methods developed for sex estimation from MSCT clinical and postmortem DICOM datasets.

	Anatomical region	Population	Sample	Image reconstruction	Reported highest and lowest accuracy	Reference
Skull	Endocranial cavity	Colombian	Clinical	Sagittal slices + 3D VRs	60%–95.8%	Isaza, Díaz, Bedoya, Monsalve, and Botella (2014)
	Exocranial surface	French	Clinical	3D	87.4%–90.3%[a]	Musilová, Dupej, Velemínská, Chaumoitre, and Bruzek (2016)
	Frontal bone	Turkish	Clinical	3D	76.3%–78.8%	Bulut, Petaros, Hizliol, Wärmländer, and Hekimoglu (2016)
	Foramen magnum	Swiss	Postmortem	Axial slices	43.4%–81.3%	Edwards, Viner, Schweitzer, and Thali (2013)
		French	Clinical	3D VRs	50%–100%	Seifert, Friedl, Chaumoitre, and Brůžek (2017)
	Internal acoustic canal meatus	N.S.	Clinical	Axial slices	N.S.	Akansel et al. (2008)
	Mastoid	Saudi	Clinical	Axial slices	59.2%–73.8%	Madadin et al. (2015)
	Bony labyrinth	Cretan	Postmortem	3D VRs	68.2%–83.7%	Osipov et al. (2013)
	Maxillary sinuses	Iraqi	Clinical	Axial and coronal slices	53.3%–74.4%	Uthman, Al-Rawi, Al-Naaimi, and Al-Timimi (2011)
		Indian	Clinical	Coronal slices	66.7%–86.7%	Prabhat et al. (2016)
		French	Clinical	3D	68%	Radulesco, Michel, Mancini, Dessi, and Adalian (2018)
	Frontal sinuses	French	Clinical	3D	72.5%	Michel et al. (2015)
	Paranasal sinuses	Egyptian	Clinical	Axial, coronal, and sagittal slices	63%–77%	Sherif, Sheta, Ibrahim, Kaka, and Henaidy (2017)
	Piriform aperture	Egyptian	Clinical	3D VRs	61.3%–86.2%	Abdelaleem, Younis, and Kader (2016)

	Bony nose shape	German and Chinese	Clinical	3D	69.1%–84.6%	Schlager and Rüdell (2015)
	Mandible	Israeli	Clinical	Axial and sagittal slices + 3D VRs	55%–95%	Sella Tunis et al. (2017)
	Palatines	Polish	Clinical	Axial, coronal, and sagittal slices	68.35%–78.37%[a]	Tomaszewska et al. (2014)
	Canines	French	Clinical	3D VRs	100%	Tardivo et al. (2011)
Vertebrae	Second cervical	Japanese	Postmortem	Axial and sagittal slices	66.1%–94.6%	Torimitsu et al. (2016b)
	12th Thoracic	Korean	Postmortem	3D VRs	62.7%–90%	Yu et al. (2008)
	First lumbar	Chinese	Clinical	3D VRs	56.4%–88.6%	Zheng et al. (2012)
Thoracic cavity	Sternum	Turkish	Clinical	Coronal and sagittal slices	56.4%–86.1%	Ekizoglu et al. (2014)
	Thorax	Spanish	Clinical	3D	N.S.	García-Martínez, Torres-Tamayo, Torres-Sanchez, García-Río, and Bastir (2016)
	Sternum + fourth rib	Turkish	Clinical	Coronal slices	60%–89.3%	Ramadan et al. (2010)
Upper extremities	Scapula	Japanese	Postmortem	3D VRs	72.5%–96.3%	Torimitsu et al. (2016a)
		Egyptian	Clinical	3D VRs	87%–95%	Paulis and Abu Samra (2015)
	Hand bones	Egyptian	Clinical	Coronal slices + 3D VRs	46.7%–100%	Eshak, Ahmed, and Abdel Gawad (2011)
	Carpals	Malaysian	Clinical	Axial slices	84.6–97.8%	Didi, Azman, and Nazri (2016)
Lower extremities	Sacrum + coccyx	Japanese	Postmortem	Axial and sagittal slices	64.3%–85.2%	Torimitsu et al. (2017)
	Femoral condyles	Korean	Postmortem	3D VRs	72.3%–94.1%	Kim, Kwak, and Han (2013)
		French	Clinical	Axial and coronal slices +3D	63.9%–78.4%	Cavaignac et al. (2016)
	Calcaneus	Turkish	Clinical	3D VRs	66.7%–100%	Ekizoglu et al. (2017)

N.S., not specified; *VRs*, volume renders; 3D, three dimensional.

[a] Highest accuracy only was reported.

morphological and size variations in skeletal elements between males and females. Through discriminate function analyses, these techniques show dimorphic differences with accuracy rates similar to many techniques established from the traditional skull and pelvic gross bones.

In addition to MSCT, cone beam CT (CBCT) and micro-CT (μCT) have also been utilized to investigate sexual dimorphism. Their applications to biological anthropology have remained limited however, as these modalities are less commonly employed in routine clinical and postmortem medical diagnoses. With CBCT, the X-rays are divergent from a single cone-shaped scanner that completes a 360-degree rotation around the stationary body part, capturing approximately 600 X-rays (Scarfe, Li, Aboelmaaty, Scott, & Farman, 2012). CBCT is typically used clinically for the examination of dentition and the maxillofacial region. Consequently, studies using CBCT have focused on recording anthropometric measurements from images to identify sexual dimorphism in the sinuses (Paknahad, Shahidi, & Zarei, 2017) and mandible (de Oliveira Gamba, Alves, & Haiter-Neto, 2014, 2016), with reported accuracies of 76%, 86.1%, and 93.33%–94.74% respectively. The principle of CBCT applies to μCT functionality; however, the cross-sections of μCT are in micrometer ranges and, therefore, scan only small anatomical structures (Boerckel, Mason, McDermott, & Alsberg, 2014). This modality has only recently been used to investigate sexual dimorphism. Kramer, Lopez-Capp, Michel-Crosato, and Biazevic (2018) used the minute bone detail provided with μCT imaging to precisely measure various dimensions of the mastoid process as a means to quantify its volume; the results of which showed to be sexually dimorphic with accuracies as high as 81.45%.

Utility of MSCT in disaster victim identification

The use of medical imaging to assist in the estimation of sex plays a valuable role in disaster victim identification (DVI). In DVI situations, human remains, either body parts or complete bodies, are typically differentially preserved and/or comingled (Blau, Robertson, & Johnstone, 2008; Brough, Morgan, & Rutty, 2015). In such cases, information pertaining to the sex of the individual plays an important role in the initial triage of the investigation prior to scientific identification (i.e., fingerprints, odontology, or deoxyribonucleic acid—DNA). For example, in the 2009 "Black Saturday" bushfires in Victoria, Australia, family members died together in their homes, which were each designated a DVI scene. Multiple bodies in a single location often resulted in comingling of the remains, which, together with exposure to the effects of fire, made identification complex. As part of the DVI triage process, all remains underwent a postmortem MSCT scan on admission to the Victorian Institute of Forensic Medicine. At one particular DVI scene, it was known that a family of five (mother, father, and three children) were together in their house. Five sets of remains were recovered; however, due to the effects of fire and comingling, the identification process was complex. MSCT scanning of one of the sets of remains showed skeletal pelvic morphology consistent with a female (Fig. 3). This information quickly supported the hypothesis that the individual was likely to be the wife/mother, which then expedited her positive identification using DNA.

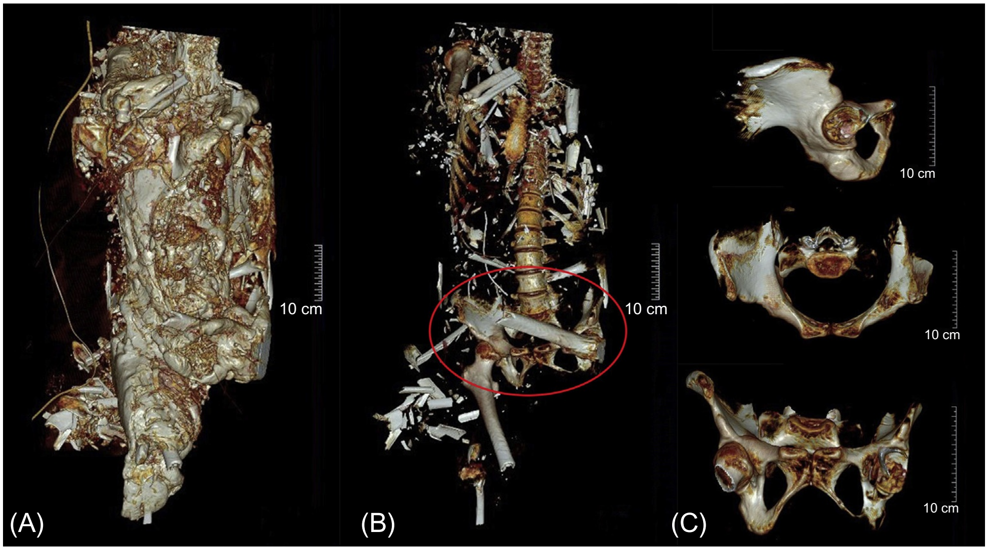


Fig. 3 Postmortem MSCT scanning of a body as part of the triage DVI process in the 2009 Victorian (Australian) bushfires. The MSCT scan was able to show that, although the remains were differentially preserved as a result of the effects of fire (A), sexually dimorphic skeletal elements (i.e., pelvis) were still present (B), and that these bones exhibited morphological features consistent with a female (C). *(Image courtesy: Victorian Institute of Forensic Medicine.)*

Utility of MSCT in archaeological contexts

In archaeological contexts, MSCT has replaced and/or augmented the traditional morphological estimations of sex from gross bones to assist with reconstructions of past life histories and possible identifications. One such example of imaging being used as a substitute for gross bone is the Sulman "princess" mummy where, due to the complexities associated with destroying ancient mummified tissues, Gardner, Garvin, Nelson, Vascotto, and Conlogue (2004) used imaging for their anthropological examination. As sex could not be morphologically estimated from conventional X-rays due to the anatomical malalignment of the innominate, Gardner et al. (2004) highlighted how advancements in imaging with the development of MSCT provided the best means to visualize the pelvic morphological features. The use of imaging allowed the researches to accurately estimate the individual's sex as female, a finding which further augmented their interpretations of the mummy's possible "royal status" (Gardner et al., 2004).

In cases where gross skeletal remains are available, MSCT has offered the opportunity to validate the traditional anthropological assessments of sex. This was the case in the identification of the skeletal remains of the historical figure King Richard III (Brough et al., 2016). In this case, Brough et al. (2016) used both traditional gross bone and MSCT images to analyze the standard morphological features of the pelvis and skull to estimate sex. The assessments were undertaken by two anthropologists who were blinded to each other's findings.

From both gross bone and MSCT the individual was estimated to be male, although, interestingly, the positioning of the bones in the CT scan hindered an assessment of sex based on the pelvis (Brough et al., 2016). Consequently, standard morphological methods for the skull only were relied upon (Brough et al., 2016).

Magnetic resonance imaging

Magnetic resonance imaging, first applied to diagnostic medicine by physician Raymond Vahan Dmadian in 1970, does not involve ionizing radiation, but rather measures the alignment of hydrogen protons (Mikla & Mila, 2014). As the water content of bone, of which hydrogen forms a component, is low at 31.81% when still organic, compared with 83.74% in some of the body's soft tissues, MRI is ideal for viewing soft tissues rather than skeletal structures (Mitchell et al., 1945).

MRI involves releasing radiofrequency pulses through a stationary body that is situated inside a strong magnetic field (Berger, 2002). The frequency pulses cause hydrogen protons in the body to slightly alter position. Once the pulses cease, the protons shift back into their natural alignment, creating radio signal "echos" from the body (Berger, 2002; Mikla & Mila, 2014). These signals are then reconstructed to create images (Berger, 2002). Like CT, MRI takes multiple images of the body in cross-sections, which are then computer-processed and stitched together to create 2D and 3D images.

Although MRI is not ideal for viewing osseous material and is, therefore, not the imaging modality preference of anthropologists, it is excellent for viewing the brain. Consequently, MRI scans of the head are often taken for clinical diagnostic purposes, and so, on occasions, anthropologists have made use of those images to develop metric sex estimation techniques from the skull. For example, Hatipoglu, Ozcan, Hatipoglu, and Yuksel (2008) and Sabancıoğulları et al. (2012) both quantified the differences in calvarial diploe thicknesses between males and females, and Chen et al. (2011) measured craniofacial soft tissue thicknesses and nasal profiles to identify differences in the sexes. These studies took advantage of the details provided for both the soft and hard tissues that only MRI affords, and introduced novel methods for investigating anthropological questions. Through using both soft and hard tissue landmarks in these methods, all studies identified dimorphic differences between sexes with males exhibiting larger sizes than females. Comparatively, Rani et al.'s (2017) study employed a method that did not take advantage of the soft tissue detail, and instead used a method for MRI images that has also been used for CT data; that is, measuring bone dimensions and volume to investigate size differences and anatomic variability of the maxillary sinuses. Their study also identified sexually dimorphic differences, with males exhibiting larger sinuses compared to females (Rani et al., 2017).

Limitations

There are a number of limitations that need to be considered when using medical imaging to investigate sexual dimorphism. First, there has been limited validation of the methods developed from images for sex estimation. As the application of medical imaging to biological anthropology is relatively recent, the methods that have been developed from medical images are still quite new and novel. As such, most methods have not yet had time to be validated for use in practice. Further, limited research has been done to investigate whether the sex estimation methods that have specifically been developed from images may be applicable for use on gross bone, and whether the methods that have been developed from gross bone may be applicable for use on medical images. Studies by Franklin et al. (2013), Lorkiewicz-Muszyńska et al. (2015), and Stull, Tise, Ali, and Fowler (2014) have found CT images of bone to generally be comparable with gross bone. This suggests that both gross bone and CT methods may be used interchangeably for estimating sex. This conclusion has also been supported by Mestekova, Bruzek, Veleminska, and Chaumoitre (2015) and Johnstone-Belford, Flavel, and Franklin (2018) who validated sex estimation methods developed from gross bone on CT data. Although there has been some work validating the results of gross bone analyses on CT, further comparison studies are still required, and validation work for comparing X-ray and MRI imaging with gross bone also needs to be undertaken before sex estimation techniques may be applied to images with confidence. Second, reported accuracy rates of the sex estimation techniques developed using images are mostly below the generally accepted accuracy threshold of 80% (Christensen, Passalacqua, & Bartelink, 2014), and especially below the considered "optimal" threshold of 95% (Krishan et al., 2016). As such, these methods should be used cautiously until further validations and refinements have, where possible, been made. Third, the practicality of applying some of these medical imaging techniques to daily biological anthropology work, either archaeological or medico-legal, is questionable. Many of the novel techniques that have been developed using CT and MRI imaging (e.g., quantifying paransal sinus volume) are only possible to use on remains that have been CT- or MRI-scanned. Furthermore, purchasing and maintaining CT and MRI equipment is costly, and so, consequently, these resources are typically only available for access from hospitals or medico-legal facilities. They also require specialist training in how to use the imaging modalities and interpret the findings, for which consultation with a radiologist is advised. Consequently, logistical drawbacks with accessing medical imaging resources will make it impractical for many biological anthropologists to use these resources in regular analyses of sex estimation from skeletal remains.

Conclusion

The application of medical imaging to the discipline of biological and forensic anthropology has revolutionized how sexual dimorphism may be investigated in the human skeleton. Since the late 1800s, X-rays, and more recently MRI and particularly CT, have been increasingly implemented in biological anthropology to advance current understandings of sexual dimorphism in the skeleton, beyond the standard anatomical structures of the pelvis, skull, and long bones. Furthermore, imaging has provided a means to investigate sexual dimorphism in modern populations and will ensure techniques developed on gross bones are able to be validated and refined. Although imaging modalities have opened new possibilities for investigating sexual dimorphism, they should not be considered superior to the traditional gross bone methods. Rather, medical images should be considered a means to augment the gross bone analysis and/or an alternative means to examine osseous tissues in cases where remains are differentially preserved. In an ever-evolving world of technology and medical advancements, medical imaging will continue to play an essential role in research and practice that will further improve the abilities of anthropologists to estimate sex from the human skeleton.

References

Abdel Fatah, E. E., Shirley, N. R., Jantz, R. L., & Mahfouz, M. R. (2014). Improving sex estimation from crania using a novel three-dimensional quantitative method. *Journal of Forensic Sciences*, *59*(3), 590–600.

Abdel Moneim, W. M., Abdel Hady, R. H., Abdel Maaboud, R. M., Fathy, H. M., & Hamed, A. M. (2008). Identification of sex depending on radiological examination of foot and patella. *The American Journal of Forensic Medicine and Pathology*, *29*(2), 136–140.

Abdelaleem, S. A., Younis, R. H. A., & Kader, M. A. (2016). Sex determination from the piriform aperture using multi slice computed tomography: Discriminant function analysis of Egyptian population in Minia Governorate. *Egyptian Journal of Forensic Sciences*, *6*, 429–434.

Aichinger, H., Dierker, J., Joite-Barfuß, S., & Säbel, M. (2012). *Radiation exposure and image quality in x-ray diagnostic radiology: Physical principles and clinical applications*. Berlin: Springer-Verlag.

Akansel, G., Inan, N., Kurtas, O., Tahsin Sarisoy, H., Arslan, A., & Demirci, A. (2008). Gender and the lateral angle of the internal acoustic canal meatus as measured on computerized tomography of the temporal bone. *Forensic Science International*, *178*(2), 93–95.

Belaldavar, C., Kotrashetti, V. S., Hallikerimath, S. R., & Kale, A. D. (2014). Assessment of frontal sinus dimensions to determine sexual dimorphism among Indian adults. *Journal of Forensic Dental Sciences*, *6*(1), 25–30.

Berger, A. (2002). Magnetic resonance imaging. *British Medical Journal*, *324*, 35.

Bierry, G., Le Minor, J. -M., & Schmittbuhl, M. (2010). Oval in males and triangular in females? A quantitative evaluation of sexual dimorphism in the human obturator foramen. *American Journal of Physical Anthropology*, *141*, 626–631.

Biwasaka, H., Aoki, Y., Sato, K., Tanijiri, T., Fujita, S., Dewa, K., et al. (2012). Analyses of sexual dimorphism of reconstructed pelvic computed tomography images of contemporary Japanese using curvature of the greater sciatic notch, pubic arch and greater pelvis. *Forensic Science International*, *219*, 288.e1–288.e8.

Blau, S., & Hill, A. J. (2014). Pediatric anthropology and odontology. In K. A. Collins & R. W. Byard (Eds.), *Forensic pathology of infancy and childhood* (pp. 1081–1125). New York: Springer.

Blau, S., Robertson, S., & Johnstone, M. (2008). Disaster victim identification: New applications for postmortem computed tomography. *Journal of Forensic Sciences*, *53*(4), 956–961.

Boerckel, J. D., Mason, D. E., McDermott, A. M., & Alsberg, E. (2014). Microcomputed tomography: Approaches and applications in bioengineering. *Stem Cell Research and Therapy*, *5*, 144.

Brogdon, B. G. (2011). Definitions in forensics and radiology. In M. J. Thali, M. D. Viner, & B. G. Brogdon (Eds.), *Brogdon's forensic radiology* (pp. 3–7). Boca Raton: CRC Press.

Brough, A., Morgan, B., Robinson, C., Appleby, J., Buckley, R., & Rutty, G. (2016). Biological profiling of Richard III using post-mortem computed tomography scanning. *Journal of Forensic Radiology and Imaging*, *5*, 31–37.

Brough, A., Morgan, B., & Rutty, G. (2015). Postmortem computed tomography (PMCT) and disaster victim identification. *La Radiologia Medica*, *120*(9), 866–873.

Bulut, O., Petaros, A., Hizliol, İ., Wärmländer, S. K. T. S., & Hekimoglu, B. (2016). Sexual dimorphism in frontal bone roundness quantified by a novel 3D-based and landmark-free method. *Forensic Science International*, *261*, 162.e161–162.e165.

Capitaneanu, C., Willems, G., Jacobs, R., Fieuws, S., & Thevissen, P. (2017). Sex estimation based on tooth measurements using panoramic radiographs. *International Journal of Legal Medicine*, *131*, 813–821.

Cavaignac, E., Savall, F., Faruch, M., Reina, N., Chiron, P., & Telmon, N. (2016). Geometric morphometric analysis reveals sexual dimorphism in the distal femur. *Forensic Science International*, *259*, 246.e1–246.e5.

Chen, F., Chen, Y., Yu, Y., Qiang, Y., Liu, M., Fulton, D., et al. (2011). Age and sex related measurement of craniofacial soft tissue thickness and nasal profile in the Chinese population. *Forensic Science International*, *212*, 272.e1–272.e6.

Chhem, R., & Brothwell, D. R. (2008). *Paleoradiology imaging: Mummies and fossils*. Berlin: Springer.

Christensen, A., Passalacqua, N., & Bartelink, E. (2014). *Forensic anthropology: Current methods and practice*. United States: Academic Press.

Christensen, A., Smith, M., Gleiber, D., Cunnignham, D., & Wescott, D. (2018). The use of X-ray computed tomography technologies in forensic anthropology. *Forensic Anthropology*, *1*(2), 124–140.

Davy-Jow, S. L., & Decker, S. J. (2014). Virtual anthropology and virtopsy in human identification. In X. Mallet, T. Blythe, & R. Berry (Eds.), *Advances in forensic human identification* (pp. 271–288). Boca Raton: CRC Press.

de Oliveira Gamba, T., Alves, M. C., & Haiter-Neto, F. (2014). Analysis of sexual dimorphism by locating the mandibular canal in images of cone-beam computed tomography. *Journal of Forensic Radiology and Imaging*, *2*, 72–76.

de Oliveira Gamba, T., Alves, M. C., & Haiter-Neto, F. (2016). Mandibular sexual dimorphism analysis in CBCT scans. *Journal of Forensic and Legal Medicine*, *38*, 106–110.

Dedouit, F., Savall, F., Mokrane, F. -Z., Rousseau, H., Crubézy, E., Rougé, D., et al. (2014). Virtual anthropology and forensic identification using multidetector CT. *British Journal of Radiology*, *87*. 20130468. https://doi.org/10.1259/bjr.20130468.

Dedouit, F., Telmon, N., Rousseau, H., Crubézy, E., Joffre, F., & Rougé, D. (2011). Modern cross-sectional imaging in anthropology. In M. J. Thali, M. D. Viner, & B. G. Brogdon (Eds.), *Brogdon's forensic radiology* (pp. 107–125). Boca Raton: CRC Press.

DeSilva, R., Flavel, A., & Franklin, D. (2014). Estimation of sex from the metric assessment of digital hand radiographs in a Western Australian population. *Forensic Science International*, *244*, 314.e1–314.e7.

Didi, A. L. M., Azman, R. R., & Nazri, M. (2016). Sex determination from carpal bone volumes: A multi detector computed tomography (MDCT) study in a Malaysian population. *Legal Medicine*, *20*, 49–52.

Edwards, K., Viner, M. D., Schweitzer, W., & Thali, M. J. (2013). Sex determination from the foramen magnum. *Journal of Forensic Radiology and Imaging*, *1*, 186–192.

Ekizoglu, O., Hocaoglu, E., Inci, E., Bilgili, M. G., Solmaz, D., Erdil, I., et al. (2014). Sex estimation from sternal measurements using multidetector computed tomography. *Medicine*, *93*(27), e240.

Ekizoglu, O., Inci, E., Palabiyik, F. B., Can, I. O., Er, A., Bozdag, M., et al. (2017). Sex estimation in a contemporary Turkish population based on CT scans of the calcaneus. *Forensic Science International*, *279*, 310.e1–310.e6.

Eshak, G. A., Ahmed, H. M., & Abdel Gawad, E. A. M. (2011). Gender determination from hand bones length and volume using multidetector computed tomography: A study in Egyptian people. *Journal of Forensic and Legal Medicine*, *18*(6), 246–252.

Franklin, D., Cardini, A., Flavel, A., Kuliukas, A., Marks, M. K., Hart, R., et al. (2013). Concordance of traditional osteometric and volume-rendered MSCT interlandmark cranial measurements. *International Journal of Legal Medicine*, *127*, 505–520.

Franklin, D., Swift, L., & Flavel, A. (2016). 'Virtual anthropology' and radiographic imaging in the forensic medical sciences. *Egyptian Journal of Forensic Sciences*, *6*, 31–43.

García-Martínez, D., Torres-Tamayo, N., Torres-Sanchez, I., García-Río, F., & Bastir, M. (2016). Morphological and functional implications of sexual dimorphism in the human skeletal thorax. *American Journal of Physical Anthropology*, *161*, 467–477.

Gardner, J., Garvin, G., Nelson, A., Vascotto, G., & Conlogue, G. (2004). Paleoradiology in mummy studies: The Sulman mummy project. *Canadian Association of Radiologists Journal*, *55*(4), 228–234.

Gonzalez, R. A. (2012). Determination of sex from juvenile crania by means of discriminant function analysis. *Journal of Forensic Sciences*, *57*(1), 24–34.

Hatipoglu, H. G., Ozcan, H. N., Hatipoglu, U. S., & Yuksel, E. (2008). Age, sex and body mass index in relation to calvarial diploe thickness and craniometric data on MRI. *Forensic Science International*, *182*, 46–51.

Henderson, C., & Cardoso, F. (2018). *Identified skeletal collections: The testing ground of anthropology?* Oxford: Archaeopress.

Holdren, J. P., & Lander, E. (2016). *Report on forensic science in criminal courts: ensuring scientific validity of feature-comparison methods*. Retrieved from the White House President Barack Obama, Office of Science and Technology Washington, President's Council of Advisors on Science and Technology. website: https://obamawhitehouse.archives.gov/administration/eop/ostp/pcast/docsreports.

Holton, N. E., Alsamawi, A., Yokley, T. R., & Froehle, A. W. (2016). The ontogeny of nasal shape: An analysis of sexual dimorphism in a longitudinal sample. *American Journal of Physical Anthropology*, *160*, 52–61.

Horbaly, H. E., Kenyhercz, M. W., Hubbe, M., & Steadman, D. W. (2019). The influence of body size on the expression of sexually dimorphic morphological traits. *Journal of Forensic Sciences*, *64*(1), 52–57.

Hsiao, T. -H., Chang, H. -P., & Liu, K. -M. (1996). Sex determination by discriminant function analysis of lateral radiographic cephalometry. *Journal of Forensic Sciences*, *41*(5), 792–795.

Isaza, J., Díaz, C. A., Bedoya, J. F., Monsalve, T., & Botella, M. C. (2014). Assessment of sex from endocranial cavity using volume-rendered CT scans in a sample from Medellín, Colombia. *Forensic Science International*, *234*, 186.e1–186.e10.

Johnstone-Belford, E., Flavel, A., & Franklin, D. (2018). Morphoscopic observations in clinical pelvic MDCT scans: Assessing the accuracy of the Phenice traits for sex estimation in a Western Australian population. *Journal of Forensic Radiology and Imaging*, *12*, 5–10.

Kim, D. -I., Kwak, D. -S., & Han, S. -H. (2013). Sex determination using discriminant analysis of the medial and lateral condyles of the femur in Koreans. *Forensic Science International*, *233*, 121–125.

Klales, A., Ousley, S., & Vollner, J. (2012). A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *American Journal of Physical Anthropology*, *149*, 104–114.

Kramer, N. A., Lopez-Capp, T. T., Michel-Crosato, E., & Biazevic, M. G. H. (2018). Sex estimation from the mastoid process using micro-CT among Brazilians: Discriminant analysis and ROC curve analysis. *Journal of Forensic Radiology and Imaging*, *14*, 1–7.

Krishan, K., Chatterjee, P. M., Kanchan, T., Kaur, S., Baryah, N., & Singh, R. K. (2016). A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. *Forensic Science International*, *261*, 165.e1–165.e8.

Lee, U. -Y., Kim, I. -B., & Kwak, D. -S. (2015). Sex determination using discriminant analysis of upper and lower extremity bones: New approach using the volume and surface area of digital model. *Forensic Science International*, *253*, 135.e1–135.e4.

Lewis, M. E. (2007). *The bioarchaeology of children: Perspectives from biological and forensic anthroplogy*. Cambridge: Cambridge University Press.

Lorkiewicz-Muszyńska, D., Kociemba, W., Sroka, A., Kulczyk, T., Żaba, C., Paprzycki, W., et al. (2015). Accuracy of the anthropometric measurements of skeletonized skulls with corresponding measurements of their 3D reconstructions obtained by CT scanning. *Anthropologischer Anzeiger*, *72*(3), 293–301.

Madadin, M., Menezes, R. G., Dhafeeri, O. A., Kharoshah, M. A., Ibrahim, R. A., Nagesh, K. R., et al. (2015). Evaluation of the mastoid triangle for determining sexual dimorphism: A Saudi population based study. *Forensic Science International*, *254*, 244.e1–244.e4.

Mendelejeff, D. (1869). On the relationship of the properties of the elements to their atomic weights. *Zeitscrift für Chemie*, *12*, 405–406.

McCormick, W. F., Stewart, J. H., & Langford, L. A. (1985). Sex determination from chest plate roentgenograms. *American Journal of Physical Anthropology*, *68*(2), 173–195.

Mestekova, S., Bruzek, J., Veleminska, J., & Chaumoitre, K. (2015). A test of the DSP sexing method on CT images from a modern French sample. *Journal of Foresnic Sciences*, *60*(5), 1295–1299.

Michel, J., Paganelli, A., Varoquaux, A., Piercecchi-Marti, M. -D., Adalian, P., Leonetti, G., et al. (2015). Determination of sex: Interest of frontal sinus 3D reconstructions. *Journal of Foresnic Sciences*, *60*(2), 269–273.

Mikla, V., & Mila, V. (2014). *Medical imaging technology*. London: Academic Press.

Mitchell, H. H., Hamilton, T. S., Steggerda, F. R., & Bean, H. W. (1945). The chemical composition of the adult human body and its bearing on the biochemistry of growth. *Journal of Biological Chemistry*, *158*, 625–637.

More, C. B., Vijayvargiya, R., & Saha, N. (2017). Morphometric analysis of mandibular ramus for sex determination on digital orthopantomogram. *Journal of Forensic Dental Sciences*, *9*(1), 1–5.

Musilová, B., Dupej, J., Velemínská, J., Chaumoitre, K., & Bruzek, J. (2016). Exocranial surfaces for sex assessment of the human cranium. *Forensic Science International*, *269*, 70–77.

Naikmasur, V. G., Shrivastava, R., & Mutalik, S. (2010). Determination of sex in South Indians and immigrant Tibetans from cephalometric analysis and discriminant functions. *Forensic Science International*, *197*, 122.e1–122.e6.

O'Donnell, A., Berry, S. D., & Edgar, H. J. H. (2017). Can cephalometrics discriminate between the sexes in a diverse juvenile sample? *Journal of Foresnic Sciences*, *62*(3), 735–740.

Oesterhelweg, L., & Thali, M. J. (2008). Experiences with virtual autopsy approach worldwide. In M. J. Thali, R. Dirnhofer, & P. Vock (Eds.), *The virtopsy approach: 3D optical and radiological scanning and reconstruction in forensic medicine* (pp. 475–477). Boca Raton: Taylor & Francis.

Osipov, B., Harvati, K., Nathena, D., Spanakis, K., Karantanas, A., & Kranioti, E. F. (2013). Sexual dimorphism of the bony labyrinth: A new age-independent method. *American Journal of Physical Anthropology*, *151*, 290–301.

Paknahad, M., Shahidi, S., & Zarei, Z. (2017). Sexual dimorphism of maxillary sinus dimensions using cone-beam computed tomography. *Journal of Foresnic Sciences*, *62*(2), 395–398.

Paulis, M. G., & Abu Samra, M. F. (2015). Estimation of sex from scapular measurements using chest CT in Egyptian population sample. *Journal of Forensic Radiology and Imaging*, *3*, 153–157.

Prabhat, M., Rai, S., Kaur, M., Prabhat, K., Bhatnagar, P., & Panjwani, S. (2016). Computed tomography based forensic gender determination by measuring the size and volume of the maxillary sinuses. *Journal of Forensic Dental Sciences*, *8*(1), 40–46.

Radulesco, T., Michel, J., Mancini, J., Dessi, P., & Adalian, P. (2018). Sex estimation from human cranium: Forensic and anthropological interest of maxillary sinus volumes. *Journal of Forensic Sciences*, *63*(3), 805–808.

Ramadan, S. U., Türkmen, N., Anıl Dolgun, N., Gökharman, D., Menezes, R. G., Kacar, M., et al. (2010). Sex determination from measurements of the sternum and fourth rib using multislice computed tomography of the chest. *Forensic Science International*, *197*, 120.e1–120.e5.

Rani, S. U., Rao, G. V., Kumar, D. R., Sravya, T., Sivaranjani, Y., & Kumar, M. P. (2017). Age and gender assessment through three-dimensional morphometric analysis of maxillary sinus using magnetic resonance imaging. *Journal of Forensic Dental Sciences*, *9*, 46. https://doi.org/10.4103/0975-1475.206481.

Rejtarová, O., Hejna, P., Rejtar, P., Bukač, J., Slížová, D., & Krs, O. (2009). Sexual dimorphism of ossified costal cartilage. Radiograph scan study on Caucasian men and women (Czech population). *Forensic Science International*, *191*, 110.e1–110.e5.

Rowbotham, S. K. (2016). Anthropological estimation of sex. In S. Blau & D. Ubelaker (Eds.), *Handbook of forensic anthropology and archaeology* (2nd ed., pp. 261–272). New York: Routledge.

Sabancıoğulları, V., Koşar, M. İ., Şalk, İ., Erdil, F. H., Öztoprak, İ., & Çimen, M. (2012). Diploe thickness and cranial dimensions in males and females in mid-Anatolian population: an MRI study. *Forensic Science International*, *219*, 289.e1–289.e7.

Sai kiran, C., Ramaswamy, P., & Khaitan, T. (2014). Frontal sinus index – a new tool for sex determination. *Journal of Forensic Radiology and Imaging*, *2*, 77–79.

Sambhana, S., Sanghvi, P., Mohammed, R. B., Shanta, P. P., Thetay, A. A. R., & Chaudhary, V. S. (2016). Assessment of sexual dimorphism using digital orthopantomographs in South Indians. *Journal of Forensic Dental Sciences*, *8*, 180. https://doi.org/10.4103/0975-1475.195113.

Scarfe, W. C., Li, Z., Aboelmaaty, W., Scott, S. A., & Farman, A. G. (2012). Maxillofacial cone beam computed tomography: Essence, elements and steps to interpretation. *Australian Dental Journal*, *57*(1 Suppl), 46–60.

Schlager, S., & Rüdell, A. (2015). Analysis of the human osseous nasal shape—Population differences and sexual dimorphism. *American Journal of Physical Anthropology*, *157*, 571–581.

Seifert, Z., Friedl, L., Chaumoitre, K., & Brůžek, J. (2017). Applicability and limitations of sex assessment based on foramen magnum. *Forensic Science International*, *271*, 126.e1–126.e9.

Sella Tunis, T., Sarig, R., Cohen, H., Medlej, B., Peled, N., & May, H. (2017). Sex estimation using computed tomography of the mandible. *International Journal of Legal Medicine*, *131*, 1691–1700.

Shehri, F. A., & Soliman, K. E. A. (2015). Determination of sex from radiographic measurements of the humerus by discriminant function analysis in Saudi population, Qassim region, KSA. *Forensic Science International*, *253*, 138.e1–138.e6.

Sherif, N. A. E. -H., Sheta, A. A. E. -M., Ibrahim, M. E., Kaka, R. A. E. -M., & Henaidy, M. F. (2017). Evaluation of the paranasal sinuses dimensions in sex estimation among a sample of adult egyptians using multidetector computed tomography. *Journal of Forensic Radiology and Imaging*, *11*, 33–39.

Sidhu, R., Chandra, S., Devi, P., Taneja, N., Sah, K., & Kaur, N. (2014). Forensic importance of maxillary sinus in gender determination: A morphometric analysis from Western Uttar Pradesh, India. *European Journal of General Dentistry*, *3*(1), 53–56.

Stull, K. E., L'Abbé, E. N., & Ousley, S. D. (2017). Subadult sex estimation from diaphyseal dimensions. *American Journal of Physical Anthropology*, *163*, 64–74.

Stull, K. E., Tise, M. L., Ali, Z., & Fowler, D. R. (2014). Accuracy and reliability of measurements obtained from computed tomography 3D volume rendered images. *Forensic Science International*, *238*, 133–140.

Tardivo, D., Sastre, J., Ruquet, M., Thollon, L., Adalian, P., Leonetti, G., et al. (2011). Three-dimensional modeling of the various volumes of canines to determine age and sex: A preliminary study. *Journal of Forensic Sciences*, *56*(3), 766–770.

The University of New Mexico. (2018). *UNM database of deceased people a national first*. https://carc.unm.edu/research/unm-database-of-deceased-people-a-national-first.html.

Tomaszewska, I. M., Frączek, P., Gomulska, M., Pliczko, M., Śliwińska, A., Sałapa, K., et al. (2014). Sex determination based on the analysis of a contemporary Polish population's palatine bones: A computed tomography study of 1,200 patients. *Folia Morphologica*, *73*(4), 462–468.

Torimitsu, S., Makino, Y., Saitoh, H., Sakuma, A., Ishii, N., Yajima, D., et al. (2016a). Sex estimation based on scapula analysis in a Japanese population using multidetector computed tomography. *Forensic Science International*, *262*, 285.e1–285.e5.

Torimitsu, S., Makino, Y., Saitoh, H., Sakuma, A., Ishii, N., Yajima, D., et al. (2016b). Sexual determination based on multidetector computed tomographic measurements of the second cervical vertebra in a contemporary Japanese population. *Forensic Science International*, *266*, 588.e1–588.e6.

Torimitsu, S., Makino, Y., Saitoh, H., Sakuma, A., Ishii, N., Yajima, D., et al. (2017). Sex determination based on sacral and coccygeal measurements using multidetector computed tomography in a contemporary Japanese population. *Journal of Forensic Radiology and Imaging*, *9*, 8–12.

Uldin, T. (2017). Virtual anthropology – A brief review of the literature and history of computed tomography. *Forensic Sciences Research*, *2*(4), 165–173.

US Supreme Court. (1993). *Daubert et ux., individually and as guardians ad litem for Dabert, et al. v. Merrell Dow Pharmaceuticals, inc. Vol, 509* (pp. 579–601). Retrieved from Justia US Supreme Court. website: https://supremejustia.com/cases/federal/us/509/579.

Uthman, A. T., Al-Rawi, N. H., Al-Naaimi, A. S., & Al-Timimi, J. F. (2011). Evaluation of maxillary sinus dimensions in gender determination using helical CT scanning. *Journal of Forensic Sciences*, *56*(2), 403–408.

Veyre-Goulet, S. A., Mercier, C., Robin, O., & Guérin, C. (2008). Recent human sexual dimorphism study using cephalometric plots on lateral teleradiography and discriminant function analysis. *Journal of Forensic Sciences*, *53*(4), 786–789.

Walker, P. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, *136*, 39–50.

Yu, S. -B., Lee, U. -Y., Kwak, D. -S., Ahn, Y. -W., Jin, C. -Z., Zhao, J., et al. (2008). Determination of sex for the 12th toracic vertebra by morphometry of three-dimensional reconstructed vertebral models. *Journal of Forensic Sciences*, *53*(3), 620–625.

Zech, W. -D., Hatch, G., Siegenthaler, L., Thali, M. J., & Lösch, S. (2012). Sex determination from os sacrum by postmortem CT. *Forensic Science International*, *221*, 39–43.

Zheng, W. X., Cheng, F. B., Cheng, K. L., Tian, Y., Lai, Y., Zhang, W. S., et al. (2012). Sex assessment using measurements of the first lumbar vertebra. *Forensic Science International*, *219*, 285.e1–285.e5.