British Journal of Pharmacology

British Journal of Pharmacology (2017) **174** 2801–2804 2801

EDITORIAL

Updating the guidelines for data transparency in the British Journal of Pharmacology – data sharing and the use of scatter plots instead of bar charts

Correspondence Amrita Ahluwalia, British Journal of Pharmacology, The Schild Plot, 16 Angel Gate, City Road, London, EC1V 2PT, UK. E-mail: info@bps.ac.uk

Christopher H George, S Clare Stanford, Steve Alexander, Giuseppe Cirino, James R Docherty, Mark A Giembycz, Daniel Hoyer, Paul A Insel, Angelo A Izzo, Yong Ji, David J MacEwan, Christopher G Sobey, Sue Wonnacott and Amrita Ahluwalia

British Journal of Pharmacology, London, UK

The oft-termed 'crisis' in reproducibility of preclinical investigations (Prinz *et al.,* 2011; Begley and Ellis, 2012) continues to grab the headlines, not only in scientific journals but also in the lay press (Guardian, 2015; Economist, 2017). Recent reports by eLife have confirmed that some attempts to reproduce 'key' cancer papers by the 'Reproducibility Project: Cancer Biology' were successful, while others were not (Aird *et al.,* 2017; Kandela *et al.,* 2017; Mantis *et al.,* 2017; Horrigan *et al.,* 2017a,b). This has added further fuel to the fire that was first stoked up by the findings from Bayer (Prinz *et al.*, 2011) and Amgen (Begley and Ellis, 2012) and supported by many others (for e.g. see McGrath and Lilley, 2015; Liu *et al.,* 2016; Ortuno *et al.,* 2016; Wang *et al.,* 2016). All of this activity has led to soul-searching within the research community, prompting a critical re-appraisal of the processes of implementation and publication of preclinical research. In particular, and relevant to this editorial, this re-appraisal has illustrated that 'replication' of preclinical research is not as simple a task as the word suggests. To replicate work, it is essential to have as much transparency, regarding the study and its results, as possible. Specifically, methods, tools, cells, animals, instruments and conditions must be described in sufficient detail. This imperative underlies the need for initiatives that improve the design, interpretation and reporting of experimental data (CAMARADES, n.d.; Kilkenny *et al.,* 2010). Journals and publishers have addressed many issues concerning the rigour and transparency of experimental design, and the British Journal of Pharmacology (BJP) is amongst these. As a starting point, the BJP has addressed some of the concerns regarding the reporting of animal experiments through adoption of the ARRIVE guidelines and the development of a series of Design and Analysis guidelines for preclinical research (Curtis *et al.,* 2015; McGrath and Lilley, 2015; McGrath *et al.*, 2015). The 18-point Declaration (see Table 1 in Curtis *et al.,* 2015) has sought to strengthen the experimental design, the conduct and the reporting of research published in the BJP.

To further improve reproducibility of research findings, the BJP has examined two aspects of data reporting that are the subject of intense debate: (1) the extent to which raw data should be made accessible to readers and (2) the format for presenting the data in a way that reveals qualities of the datasets that underpin the validity of authors' conclusions. Proceeding in parallel with this debate has been the stipulation by an increasing number of research councils and granting agencies that researchers receiving grants must comply with the FAIR initiative (i.e. that data should be **F**indable, **A**ccessible, **I**nteroperable and **R**euseable) (see Data handling and sharing policies of the following: EC, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf; MRC, https://www.mrc.ac.uk/research/policies-and-guidance-for-researchers/data-sharing/; NIH, https://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf; and Wellcome Trust, https://wellcome.ac.uk/what-we-do/our-work/open-research). The SHERPA/JULIET database of funders' research data policies shows that 42/60

DOI:10.1111/bph.13925

(70%) now require authors to comply with their policy on accessible data. Here, we outline our views on how the BJP should respond to this debate.

Data sharing

It is self-evident that any practice that increases transparency, rigour and accessibility of data will benefit both expert and non-expert communities and should help mitigate the failures of reproducibility. However, the practicalities of data sharing are confusing and complex and the relative merits of freely-accessible data sharing, versus sharing on request, are unresolved. Moreover, what 'data sharing' means is often not explicitly defined: for example, how 'raw' should the data be? Another problem is that standardization of data formatting and structuring will play a critical role in rendering data useful, but barriers exist to achieving this. The successes in making available DNA sequences (e.g. GenBank and dbEST) and protein structures (e.g. PDB), for which the data lend themselves to standardized structuring and phylogenic profiling, will be difficult to replicate in other types of datasets: for example, the minute-or-so cellular patch-clamp traces acquired under a variety of experimental conditions or for that matter a complete set of 24 h sleep EEG (polysomnography) recordings. Also, how might researchers be expected to accurately annotate and report the multitude of difficult-to-define determinants that contribute to particular experimental outcomes (e.g. the 'nuisance variables' (Button *et al.,* 2013; Krzywinski and Altman, 2013; Voelkl and Wurbel, 2016))? If a data sharing policy is to be of use, criteria must be much more explicitly defined than at present.

The natural extension to the use of Supplementary Data that often accompanies the published articles is to use digital repositories to archive and openly share research datasets (e.g. Open Science Framework, Open Microscopy, Figshare and Dryad; see Dryad, http://datadryad.org/; Figshare, https://figshare.com/; OpenMicroscopy, https://www.openmicroscopy.org/site; and Open Science Framework, https://http://www.dataone.org/software-tools/open-science-framework). However, technological platforms that enable data sharing are not yet fully developed and few low- or no-cost repositories have been set up to make available the terabytes of data typically generated by contemporary platform technologies (e.g. high-throughput imaging systems). Making data available in accessible formats (e.g. those not requiring proprietary software files) poses a problem and, although journals should not be held accountable for ensuring that the data underpinning published content is shared, identifying which party carries this responsibility (authors, funders or publishers) remains a challenge.

The editors of the BJP acknowledge that making available integrated platforms that link published data to the original component datasets for many types of common pharmacological data is presently not feasible. Such datasets include, for instance, 'raw' traces of electrophysiological measurements, large imaging files and the reams of continuous telemetry recordings. For this reason, the BJP encourages but does not mandate data sharing (i.e. at present BJP does not insist upon data sharing for publication). The BJP is actively investigating approaches to address these issues. However, we do not envisage that an appropriate repository/platform will become available in the near future.

Data presentation

Meanwhile, improvements in standards of data presentation and accessibility present a more immediately tractable issue to enhance the information in the 'two-dimensional' format of a research paper. In order to prepare publications, authors distil carefully compiled observations and readouts from multiple technical platforms into elements presented in tabulated or graphical form. The Editors of the BJP share concerns that this compaction may result in important features of the dataset being masked, or lost altogether (Drummond and Vowler, 2011; Weissgerber *et al.,* 2015). Bar charts, typically of grouped data presented as means with a descriptor of experimental error, are the most common form of graphical visualization in manuscripts submitted to this Journal and are used to present results from diverse types of experiments, including measurements on humans, *in vivo* and *ex vivo* data from studies with animals, *in vitro* studies in tissues and cell lines and from the biochemical assessment of samples (e.g. immunoblotting and RT-PCR). An illustrative example of a comparison of cell lines is described in Figure 1, which shows that bar charts do not give the reader adequate information on the variability and distribution of each sampled '*n*'. This is because bar charts frequently do not adequately convey major features of the dataset. As explained below, Figure 1 illustrates why moving away from using bar charts to visualize the entire dataset is a necessary refinement that can increase the transparency and reporting of data.

The immediate conclusions that could be drawn from the data presented using bar charts in (A) are (1) that cell lines a, b and c exhibit identical mean values of receptor activation under baseline conditions (55 units); (2) there is negligible inter-population variation (inter-group Kruskal–Wallis statistic $P > 0.9999$); and (3) the drug has no effect in any cell line. Close scrutiny of the error values may intuitively point to an increasing level of intra-group variability (a vs. b vs. c under baseline and drug-stimulated conditions), but plotting the data as a bar graph (A) masks the fact that the identical mean values of receptor activation in cell populations a, b and c are derived from values that differ considerably with respect to their ranges.

By plotting each individual '*n*' in grouped scatter plots (B), one sees that under baseline conditions, receptor activation in cell line 'a' is relatively homogeneous, in contrast with the broad normal distribution of activation in cell line 'b' and the two entirely distinct sub-populations of cell line 'c'. Note that in both (A) and (B), the identical SEM values before and after drug addition might (erroneously) suggest a highly uniform response of each cell line to the drug.

Presenting these data as scatter plots of paired measurements before and after the addition of drug (C) reveals very different responses. Visualizing the data in this form leads one to conclude that the addition of drug has no effect in cell line 'a' or cell line 'b'. The data also corroborate the conclusion that the level of receptor activation under baseline conditions in sub-populations has no bearing on the response to the drug of those sub-populations. By contrast, the

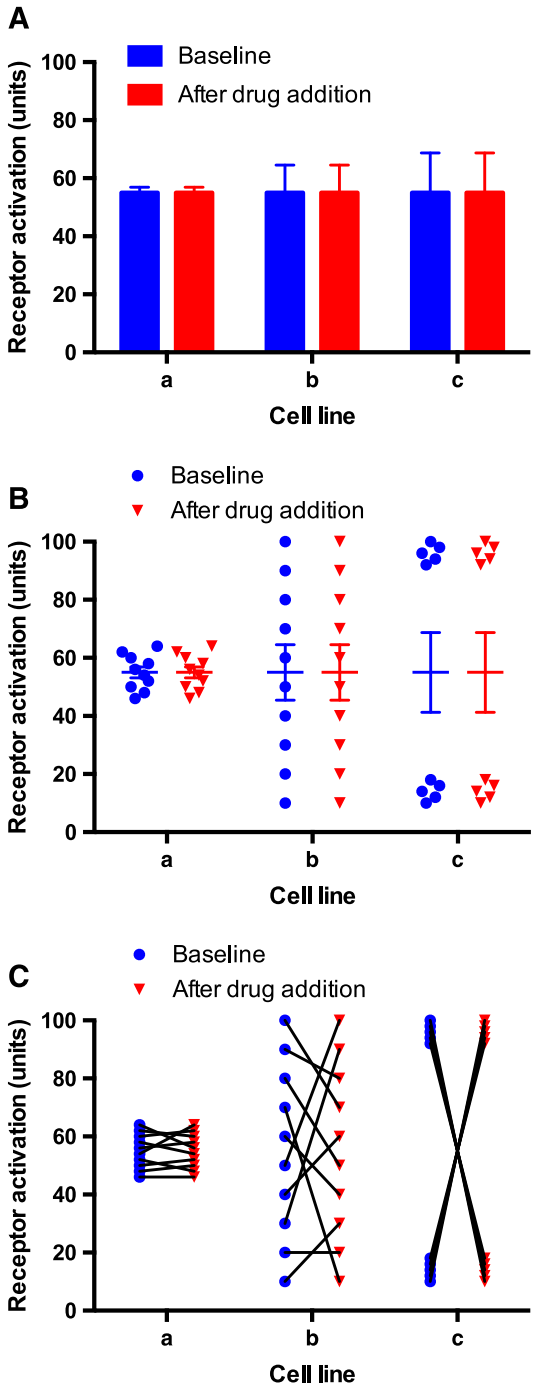

Figure 1

The extent of activation of a receptor in three cell lines a, b and c under baseline (drug-naïve) conditions and following the addition of a drug is given in arbitrary units. The same data sets are presented in three different ways: (A) bar chart, (B) grouped column scatter plot with means and error and (C) before–after scatter plot. $n = 10$ (i.e. biological replicates and not technical replicates). In this example, error bars represent the SEM although authors should consider the sampling size and distribution of '*n*' when choosing the most appropriate way of showing experimental error [e.g. SD or confidence interval (Drummond and Vowler, 2011)].

response to the drug by sub-populations of cell line 'c' is large and depends on the extent of baseline receptor activation. This type of presentation is also valid (and useful) for data derived from human and animal experiments.

Given these issues, the Editors of the BJP will now require that, where possible, numerical data (whether categorical or continuous), particularly involving two sets or paired data, should be presented using scatter-plots, before-after graphs, and other forms in which each individual '*n*' value is individually plotted, rather than using bar charts. Authors presenting data as bar charts should state that a scatter plot or before–after charts did not reveal unusual or interesting aspects of the data not obvious from the bar chart. We will update our Declaration with its checklist to show this change.

Acknowledgements

The article originated from discussions at the regular meetings of the Editor-In-Chief and the Senior Editors of BJP during 2016. C.H.G. coordinated the writing of the manuscript with contributions and edits from all the authors.

References

Aird F, Kandela I, Mantis C, Iorns E, Denis A, Williams S *et al.* (2017). Study 19: replication of Delmore *et al.*, 2001 (Cell). Elife. https://doi.org/10.17605/OSF.IO/7ZQXP.

Begley CG, Ellis LM (2012). Raise standards for preclinical cancer research. Nature 483: 531–533.

Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ *et al.* (2013). Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 14: 365–376.

CAMARADES. Available at: http://www.dcn.ed.ac.uk/camarades/default.htm (accessed 20/7/2017).

Curtis MJ, Bond RA, Spina D, Ahluwalia A, Alexander SPA, Giembycz MA *et al.* (2015). Experimental design and analysis and their reporting: new guidance for publication in BJP. Br J Pharmacol 172: 3461–3471.

Drummond GB, Vowler SL (2011). Show the data, don't conceal them. Br J Pharmacol 163: 208–210.

Economist. (2017). Available at: http://www.economist.com/blogs/economist-explains/2017/03/economist-explains-23 (accessed 20/7/2017).

Guardian. (2015). Available at: https://www.theguardian.com/commentisfree/2015/aug/28/psychology-experiments-failing-replication-test-findings-science (accessed 20/7/2017).

Horrigan S, Courville P, Sampey D, Zhou F, Cai S, Iorns E *et al.* (2017a). Study 44: replication of Berger *et al.*, 2012 (Nature). Elife. https://doi.org/10.17605/OSF.IO/JVPNW.

Horrigan S, Iorns E, Williams S, Perfito N, Errington T (2017b). Study 39: replication of Willingham *et al.*, 2012 (PNAS). Elife. https://doi.org/10.17605/OSF.IO/9PBOS.

Kandela I, Aird F, Iorns E, Williams S, Perfito N, Errington T (2017). Study 21: replication of Sirota *et al.*, 2011 (Science Translational Medicine). Elife. https://doi.org/10.17605/OSF.IO/HXRMM.

Kilkenny C, Browne W, Cuthill IC, Emerson M, Altman DG (2010). Animal research: reporting *in vivo* experiments: the ARRIVE guidelines. Br J Pharmacol 160: 1577–1579.

Krzywinski M, Altman N (2013). Power and sample size. Nat Methods 10: 1139–1140.

Liu D, Tseng M, Epstein LF, Green L, Chan B, Soriano B *et al*. (2016). Evaluation of recombinant monoclonal antibody SVmab1 binding to $Na_v1.7$ target sequences and block of human $Na_v1.7$ currents. F1000Research 5: 2764.

Mantis C, Kandela I, Aird F, Iorns E, Denis A, Williams S *et al*. (2017). Study 15: replication of Sugahara *et al*., 2010 (Science). Elife. https://doi.org/10.17605/OSF.IO/XU1G2.

McGrath JC, Lilley E (2015). Implementing guidelines on reporting research using animals (ARRIVE etc.): new requirements for publication in BJP. Br J Pharmacol 172: 3189–3193.

McGrath JC, McLachlan EM, Zeller R (2015). Transparency in research involving animals: the Basel Declaration and new principles for reporting research in BJP manuscripts. Br J Pharmacol 172: 2427–2432.

Ortuno D, Carlisle HJ, Miller S (2016). Does inactivation of USP14 enhance degradation of proteasomal substrates that are associated with neurodegenerative diseases? F1000Research 5: 137.

Prinz F, Schlange T, Asadullah K (2011). Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Disc 10: 712.

Voelkl B, Wurbel H (2016). Reproducibility crisis: are we ignoring reaction norms? Trends Pharm Sci 37: 509–510.

Wang S, Wen P, Wood S (2016). Effect of LXR/RXR agonism on brain and CSF Aβ40 levels in rats. F1000Research 5: 138.

Weissgerber TL, Milic NM, Winham SJ, Garovic VD (2015). Beyond bar and line graphs: time for a new data presentation paradigm. PLoS Biol 13: e1002128.