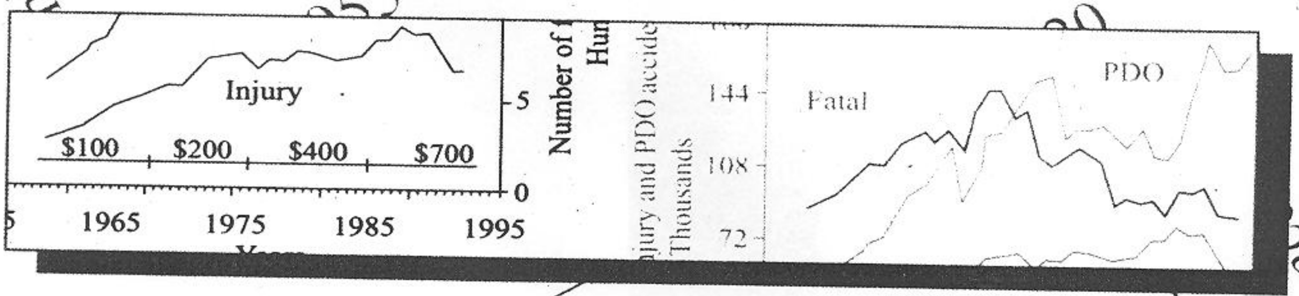
OBSERVATIONAL BEFORE–AFTER STUDIES IN ROAD SAFETY



Estimating the effect of highway and traffic engineering measures on road safety

Ezra Hauer

Pergamon

OBSERVATIONAL BEFORE—AFTER STUDIES IN ROAD SAFETY

Related Pergamon books

LESORT — Transportation and Traffic Theory

ROTHENGATTER & CARBONELL — Traffic and Transport Psychology: Theory and Application

Related Pergamon journals

Accident Analysis & Prevention
Editor: Frank Haight

Journal of Safety Research
Editor: Thomas W. Planek

Safety Science
Editor: Andrew Hale

Transportation Research Part A: Policy and Practice
Editor: Frank Haight

Transportation Research Part B: Methodological
Editor: Frank Haight

Transportation Research Part C: Emerging Technologies
Editor: Stephen Ritchie

Transport Policy
Editor: P B Goodwin

Free specimen copies available on request

OBSERVATIONAL BEFORE—AFTER STUDIES IN ROAD SAFETY

ESTIMATING THE EFFECT OF HIGHWAY AND TRAFFIC ENGINEERING MEASURES ON ROAD SAFETY

By

EZRA HAUER

Department of Civil Engineering
University of Toronto

PERGAMON
An Imprint of Elsevier Science
Amsterdam-Boston-London-New York-Oxford-Paris
San Diego-San Francisco-Singapore-Sydney-Tokyo

ELSEVIER SCIENCE Ltd
The Boulevard, Langford Lane
Kidlington, Oxford OX5 1GB, UK

First edition: 1997
Second impression: 2002
Transferred to digital printing, 2005

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: 0 08 043053 8

⊗ The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).

Printed & bound by Antony Rowe Ltd, Eastbourne

To Shoshka

CONTENTS

PREFACE

Some believe, naively, that one can know the safety effect of a treatment (countermeasure, design feature change etc.) by comparing accident rates 'before' and 'after' implementation. They are wrong. Others believe, perhaps too dogmatically, that unless one can conduct randomized experiments, the safety effect of a treatment can never be known. I hope they are wrong. Both beliefs obstruct progress. The first by polluting the store of professional knowledge by misinformation; the second by denying that one can extract defensible information from the rich professional experience that almost never involves treatment applied to randomly selected entities. In this monograph I am trying to give voice to two beliefs. First that it is possible to learn from experience even if it does not conform to the strictures of randomized experiments. Second that doing so requires a great deal of care and expertise.

Work on this monograph is a reflection of years of study. I am indebted to the National Science and Engineering Research Council of Canada, to Transport Canada - Road Safety, and to the US. Federal Highway Administration, for consistently supporting my research over the years. Draft versions went to my colleagues Rune Elvik, Per Gårder and Geoffrey Maycock. Their comments and, in particular, Maycock's interlinear led to clarification and improvement. Of course, all the remaining obscurities and deficiencies are mine. The manuscript also benefited from a detailed review by Abraham Mensah and Matt Roorda. Judging by the errors and inconsistencies they found, I fear that more may exists. I apologize to the reader for the frustration this may cause.

Ezra Hauer, January 1996.

GLOSSARY

Notation and operators:

E{ }	Expected value of the random variable inside { }.
VAR{ }	Variance of random variable inside { }.
f(.)	Function giving the expected number of accidents. The argument of the function is traffic flow and perhaps additional variables.
g(.)	Gamma probability density function.
m{ }	Sample mean of random variable inside { }.
s^2{ }	Sample variance of random variable inside { }.
v{ }	Coefficient of variation of the random variable inside { }, the standard deviation divided by the mean.
σ{ }	Standard deviation of random variable inside { }.
^	Estimate of the value under ^.
Σ	Summation from 1 to n.

Latin Symbols:

A	Traffic flow in the 'after' period.
B	Traffic flow in the 'before' period.
$C_{j,y}$	Model equation for entity j in year y divided by model equation for year 1.
F	Traffic Flow.
K(j)	Count of target accidents during the 'before' period on the treated entity j.
K	Count of target accidents during the 'before' period on the 'treatment group' of entities. The 'treatment group' is the composite entity made up of entities 1, 2, . . ., j, . . . , n.
L(j)	Count of target accidents during the 'after' period on the treated entity j.
L	Count of target accidents during the 'after' period on the 'treatment group'. The 'treatment group' is the composite entity made up of entities 1,2, . . . , j ,. . . , n.
M(j)	Count of accidents during the 'before' period on entity j of the comparison group.
M	Count of accidents during the 'before' period on the comparison entities.
N(j)	Count of accidents during the 'after' period on entity j of the comparison group.
N	Count of accidents during the 'after' period on the comparison entities.

R	Road sections are numbered 1, 2, 3, . . . ,R. Thus there are R sections in the data set and R is the label of the last section.
Y	Years in the 'before' period, numbered 1, 2, 3, . . . y, . . . , Y. Thus there are Y years in the 'before' period and Y is the last year prior to the treatment.
Z	Years in the 'after' period.
a,b	Parameters of the Gamma probability density function.
c_A	$\partial f(A)/\partial A$
c_B	$\partial f(B)/\partial B$
d, d_i	Length of road section (I).
j	Counting index that designates the j-th entity.
b	Constant used in the relationship $VAR\{\kappa\} = [E\{\kappa\}]^2/b$.
m	The number of entities in a comparison group or reference population.
n	The number of entities on which some treatment has been implemented or which each have a specific comparison group.
n(K)	Number of entities on which the accident count was K.
o	Sample odds ratio≐(N/M)/(L/K).
r_C	Comparison ratio≐(expected number of 'after' target accidents per unit of time in the comparison group)/(expected number of 'before' target accidents per unit of time in the comparison group). The subscript C in the ratio r_c stands for 'comparison group'.
r_d	Duration of 'after' period/duration of 'before' period
r_T	$\doteq \pi/E\{K\}$= (expected number of 'after' target accidents per unit of time in the treatment group)/(expected number of 'before' target accidents per unit of time in the treatment group). The subscript T in the ratio r_T stands for 'treatment group'.
$r_d(j)$	(duration of after period)/(duration of before period) for entity j. The subscript d stands for 'duration.'
$r_{tf}(j)$	(expected number of target accidents per unit of time with 'after' traffic)/(expected number of target accidents per unit of time with 'before' traffic) for entity j. The subscript t in the ratio r_t stands for 'traffic'.
t	Time.
v{.}	Ratio of the standard deviation of a random variable and its mean. This is called the 'coefficient of variation'.
y	Counter for 'years'.

Greek Symbols:

α	A number between 0 and 1 which specifies what proportion of $E\{\kappa\}$ to use in estimating κ.
α,β,γ . . .	Parameters of functions (such as, eg., when f(flow) is of the form $\alpha(\text{flow})^\beta$).

δ	$\doteq\pi-\lambda$, the reduction in the expected frequency of target accidents (by kind or consequence) per unit of time.
θ	$\doteq\lambda/\pi$, the 'index of effectiveness'.
θ^*	Approximately unbiased estimator of θ.
$\theta(j)$	Index of effectiveness for entity j=1,2,...,n.
$\bar{\theta}$	Sample mean of $\theta(j)$, j=1,2,..,n where n is the number of treated entities.
κ	E{K}, expected number of accidents of some kind occurring on a group of entities during a specified time period. Often these are target accidents and the group is the 'treatment group' of entities. At times we speak of κ in year y and this is denoted as κ_y. When we speak of κ for entity j in year y, $\kappa_{j,y}$ is used. As κ is an expected number of occurrences in a specified period of time, it is an 'expected frequency'.
$\lambda(j)$	Safety of a treated entity j in the after period.
λ	$\doteq\Sigma\lambda(I)$
μ	E{M}, expected number of accidents per unit of time on the 'comparison group of entities in the 'before' period.
ν	E{N}, expected number of accidents per unit of time on the 'comparison group of entities in the 'after' period.
$\pi(j)$	What the safety of the treated entity (or group of entities) would have been in the after period had treatment not been applied.
π	$\doteq\Sigma\pi(j)$
ω	$\doteq r_c/r_t$, the odds ratio. It is defined as $(\nu/\mu)/(\pi/\kappa)$ when treatment has been applied and by $(\nu/\mu)/(\lambda/\kappa)$ when treatment has not been applied.

Acronyms:

AADT	Annual average daily traffic
C-G method	Comparison-Group method. When accidents on a comparison group are used to account for change in causal factors
DRL	Daytime Running Lights
ITE	Institute of Transportation Engineers
MUTCD	Manual on Uniform Traffic Control Devices
RTM	Regression-to-mean
PRP	Police reporting period
VMT	Vehicle-miles-of-travel, also stands for vehicle-kilometres-of travel. If, e.g., a road section is 8 km long and carries 2000 vehicles per day, the daily VMT is 8×2000=16,000 vehicle kilometres.
vpd	Vehicles per day

Terms and Definitions:

all-way STOP control	When all intersection approaches have STOP signs and the first driver to stop has the right of way.
covariates	Same as 'predictor variables' or 'independent variables' are the known or measured traits which feature in the right hand side of a model equation.
bells & flashers	An auditory and visual warning device at rail highway grade crossings that is activated by the approach of a train.
crossbucks	A sign warning of the presence of a rail track crossing a highway.
gates	A barrier across the road lane that is activated by the approach of a train.
Poisson distribution	By the Poisson distribution, the probability to obtain an accident count L (L is a non-negative integer) is given by $\lambda^L e^{-\lambda}/L!$. In this expression λ is the mean (or expected) accident count.
safety	the expected number of accidents or of accident consequences (perhaps classified by kind or severity) occurring on an entity per unit of time during a specified time period.
two-way STOP control	When two intersection approaches (usually those with the lesser traffic flows) have STOP signs.
YIELD sign	A traffic control sign equivalent to a GIVE WAY sign.

CHAPTER 1

INTRODUCTION[1]

Transportation professionals and others make decisions which affect people's safety. When a road is built to a certain grade or when STOP signs are replaced by signals, the frequency and severity of accidents are affected. Professionalism requires that the safety consequences of such decisions be known.

Factual knowledge of this sort is not easy to come by. There is a widespread belief that personal experience as a driver, pedestrian or engineer is a good source of factual knowledge about road safety. This is dangerously untrue. Just as the experience of one person or one physician is of no use for determining the effect of smoking on the incidence of lung cancer, so the experience of a driver or a traffic engineer is insufficient to know the safety effect of paving shoulders or widening lanes. The field of road safety abounds with strongly held but unfounded opinions. Thus, most road users believe that traffic signals enhance safety; they go by their gut feeling. Most traffic engineers in North America believe that four-way STOPs[2] enhance safety only where warranted by the Manual of Uniform Traffic Control Devices[3] (FHWA, 1988); they go by professional folklore. In road safety gut feeling and folklore are frequently wrong.

One of the main sources of factual knowledge about the effect of highway and traffic engineering measures on safety is the 'observational Before-After study'. The word 'observational' may mislead; it is not intended to conjure the image of observers in the field. It is there to distinguish between an experiment that is deliberately designed to answer a question, and between the more passive pursuit of observing or noting the safety consequences of some treatment or intervention that has been implemented for purposes other than that of answering a research question. Observational studies[a] are a very imperfect source of knowledge as we shall see. Progress would be much faster if it were possible to experiment as in a laboratory; if one could change one

---

[1] Footnotes will be designated by numerical superscripts.

[2] A common North American way to control traffic at low volume intersections whereby a STOP sign is placed on all (four) approaches.

[3] A document guiding U.S. practice on the application of traffic control devices. It is commonly referred to as the MUTCD.

thing while keeping everything else constant. Barring that, one would still learn efficiently if it was possible to decide which entity (intersection, road section, driver) is to be treated and which not, by drawing numbers out of a hat. This would be a 'randomized Before-After experiment'. However, ordinarily neither option is open to us. The usual circumstance is that some treatment has been implemented on a group of entities and the task is to estimate the safety effect thereof. No randomization or experimental design are involved. Therefore, one has to make efficient use of the old workhorse - the observational Before-After study[1].

The observational Before-After study has acquired a bad name for a variety of reasons. Perhaps most important is the view of many prominent statisticians (such as Sir R.A. Fisher) according to whom statistical methods cannot be used to extract defensible conclusions from observational studies, only from randomized experiments. This attitude flies in the face of much current statistical practice. One only has to open any journal on epidemiology, econometrics, social and behavioral sciences, or engineering, to see that many (perhaps most) inferences about cause and effect are based on data that come from observational studies, not experiments. The other reason why the observational Before-After study is in disrepute is that among transportation engineers the term is often synonymous with a naive comparison of 'before' and 'after' accidents without any attempt to account for what is due to change in factors other than the treatment. In this monograph a broader interpretation is given to the term. It will encompass all techniques by which one may study the safety effect of some change that has been implemented on a group of entities (road sections, intersections, drivers, vehicles, neighborhoods, etc.). Thus, when comparison groups[2] are used, when corrections for change in conditions are applied, when time-series methods are used to make projections, in all these cases one is doing an observational Before-After study. Even if its good name cannot be entirely restored, I hope to salvage that part of its reputation that has been damaged undeservedly.

In contrast to the Before-After study, there is the Cross-Section study[3]. The latter arises when one is comparing the safety of one group of entities having some common feature (say, STOP

---

[1] That experiments in road safety are so rare is often ascribed to the unacceptability of putting people to risk. This is a leaky argument. For one, in medicine, when certain conditions are met, it seems perfectly acceptable to embark on clinical trials with dangers to health. For two, putting into use hardware such as concrete median barriers without experimentation puts people to risk; many more people than a properly designed experiment would.

[2] When prior to implementation a set of candidate entities is randomly divided into those to be treated and those not, the untreated entities form a '*control group*'. When after implementation one identifies a group of entities that have been left untreated we speak of a '*comparison group*'.

[3] This type is often referred to as the with-without or the case-control study.

controlled intersections) to the safety of a different group of entities not having that feature (say, YIELD[1] controlled intersections), in order to assess the safety effect of that feature (STOP versus YIELD sign). This is not in the domain of a Before-After study. In such a case one does not study the change of safety from 'before' to 'after' of the same entities. Rather, one is comparing the safety of two different groups of entities. Figure 1.1 is an attempt to clarify these distinctions.

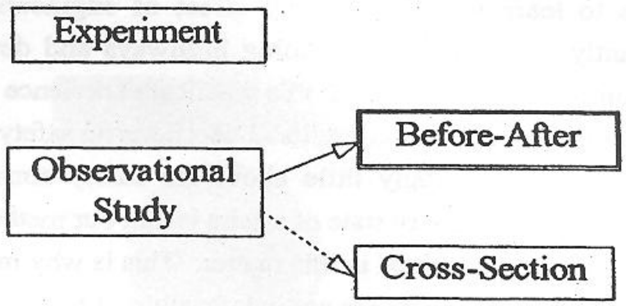


Figure 1.1. Types of study.

The natural domain of an observational Before-After study is the circumstance when the entities that are changed by the treatment retain much of their original attributes. Thus, e.g., the replacement of STOP by YIELD signs leaves the intersection geometry and setting unchanged. Similarly, the introduction of a seat belt law does not modify drivers, their travel patterns, the vehicle performance or the road network. In contrast, the natural domain of cross-section studies is the circumstance when the treatment substantially alters the entity. Thus, e.g., if a rural two lane road is to be rebuilt into a four-lane divided road with a substantially modified alignment, one might perhaps learn most about the potential safety consequences from a comparison of the safety performance of two-lane roads and four-lane divided roads. The interpretation of cross-section data will not be discussed here[2]. This monograph is devoted to the question how to conduct and interpret observational Before-After studies, the double-framed box in Figure 1.1.

---

[1] This is the North American term for GIVE WAY signs.

[2] I think that when one asks about the safety effect of treatments that leave most attributes of entities unchanged, the problems which beset the interpretation of data from observational Before-After studies are smaller and seem more surmountable than are the difficulties of defending inferences based on cross-section data. George Box (1966) said that: "The only way to find out what will happen when a complex system is disturbed is to disturb that system, not merely to observe it passively."

The audience for which this monograph is written is broader than the audience of experts in evaluative research. It includes not only those who are occasionally called upon to plan and carry out observational Before-After studies but also those who need to assess results obtained by others and to understand what they mean. The breadth of the audience creates the usual problem of style. Most of the important messages are in plain English. I hope that those who will find the mathematical passages forbidding can, by judicious skipping, still extract what is most useful.

Opportunities to learn about the safety effect of engineering decisions are abundant. Engineers are constantly building and rebuilding highways and devising schemes for signing, marking and traffic control devices. From all this practical experience one could squeeze out much factual knowledge about the impact of professional decisions on safety. Alas, we have not done so. As a result, we know embarrassingly little about the safety consequences of our actions as professionals. One reason for this sorry state of affairs is that our methodological arsenal is less than perfect and that engineers are not trained in this matter. This is why many honest and costly efforts are often inconclusive and why their results are indefensible. I hope that this monograph will be of help. There are several other reasons for the prevailing reign of ignorance. First, unlike many other engineering failures, failures of safety are not self-evident. It is difficult to pin a large accident frequency on deficient design or traffic control. Second, there is a risk that an evaluation will find an implemented measure ineffective. This makes for an inclination not to evaluate, or to do so 'in house', and not to publish what is unfavorable. These and other impediments to progress are discussed in Hauer (1988).

Perhaps more important than guidance on 'how to do' is guidance on 'what does it mean'. Professional literature abounds with respectable-looking studies. At times the conclusions are in accord with common sense, at times they look suspicious or contrary to current practice. The professional has to form an opinion about the soundness of such conclusions and the validity of the existing lore or practice. This judgement has to rest on the kind of understanding which this monograph hopes to provide[b].

The monograph is organized into three parts. The **first part** contains information that I consider **essential** both for forming an opinion about results obtained by others, as well as for the planning and analysis of an observational Before-After study. Fortunately, what is really important can be written in language accessible to all. The **second part** attempts to adapt **conventional approaches** (such as Before-After-with-Comparison-Group) to the realities of observational studies. For this, neither algebra nor statistics can be avoided. Even so, I hope that the many examples make the main lessons to be learned sufficiently transparent. The study of conventional approaches is important because much of what has been done, and much that will be done in the future, fits and will fit this mold. But it is my opinion that the conventional approaches which we inherited from those who can do randomized statistical experiments, are ill suited to the realities of road safety.

What we do in road safety are observational studies, not statistical experiments, and the means of best interpreting the two kinds are different. Accordingly, the **third part** of this monograph is devoted to an exposition of **new approaches** to the interpretation of observational Before-After studies.

Soundness of method is important. I hope that this monograph will help professionals to both understand and carry out observational Before-After studies. But progress toward knowledge-based road safety engineering requires more than soundness of method. The profession and its institutions must ensure that the opportunity to learn from experience is consistently seized, expertly exploited, dispassionately reported and carefully scrutinized. While a sound methodology is a necessary condition, it is not a sufficient one. What is needed most is a collective will of the profession to live up to its responsibilities in managing road safety.

Suggested Readings.

1. Campbell, D.T., and Stanley, J.C. (1966). Experimental and quasi-experimental designs for research. Rand McNally College Publishing Company. Chicago. Also, Cook, T.D. and Campbell, D.T. (1979) Quasi-experimentation, Design and analysis issues for field settings. Rand McNally, Chicago and Houghton Mifflin, Boston.

2. Cochran, W.G. (1983). Planning and analysis of observational studies. Edited by Moses, L.E. and Mosteller, F., John Wiley & Sons. Also, Rosenbaum, P.R., (1995). Observational studies. Springer Verlag. New York.

3. Fisher, R.A. (1971). The design of experiments. Hafner Publishing Company, New York. First published in 1935.

4. Hauer, E. (1988) A case for science-based road safety design and management. In Stammer, R.E., (ed.), Highway Safety: At the crossroads. American Society of Civil Engineers, New York.

5. Spirtes, P., Glymour, C., and Scheines, R. (1993). Causation, Prediction and Search. Springer-Verlag Inc., New York.

Endnotes.

Endnote a. I adopted the term 'observational studies' from Cochran (1983). These, he says, have two characteristics:

"1. The objective is to study the causal effects of certain agents, procedures, treatments, or programs.

2. For one reason or another, the investigator cannot use controlled experimentation, that is, the investigator cannot impose on a subject, or withhold from the subject, a procedure or treatment whose effects he desires to discover, or cannot assign subjects at random to different procedures." (p.1).

Compared with the vast literature on the design of experiments, comparatively little has been written about observational studies. Cochran's book (posthumously edited), Campbell & Stanley (1966), Cook & Campbell (1979), and Rosenbaum (1995), are the books I know. This imbalance is symptomatic of a peculiar cleavage in the collective psyche of statisticians. As Spirtes et al. point out (1993, preface), " . . . statistical methods are routinely used to justify causal inferences from data not obtained from randomized experiments, and sample statistics are used to predict the effect of policies, manipulations and experiments. Without these uses the profession of statistics would be a much smaller business. (Yet) . . . the discipline thriving from such uses assures its audience that they are unwarranted." It only remains to add that, because of the emphasis on the orthodoxy of randomized experiments, the statistical methods used to extract information from observational studies tend to be those developed for and taught to experimenters. This is a crutch but also a straitjacket. Method should fit reality, reality should not be twisted to suit the method. Part III of this monograph addresses this issue head-on.

Endnote b. "I am very sorry, Pyrophilus, that to the many (elsewhere enumerated) difficulties which you may meet with, and must therefore surmount, in the serious and effectual prosecution of experimental philosophy I must add one discouragement more, which will perhaps as much surprise as dishearten you; and it is, that besides that you will find (as we elsewhere mention) many of the experiments published by authors, or related to you by persons you converse with, false and unsuccessful (besides this I say), you will meet with several observations and experiments which, though communicated for true by candid authors or undistrusted eye witnesses, or perhaps recommended by your own experience, may, upon further trial, disappoint your expectation, either not at all succeeding constantly, or at least varying much from what you expected." Robert Boyle, 1673, Concerning the Unsuccessfulness of Experiments. Quoted by R.A. Fisher in the Design of Experiments. 1959.

PART I
ESSENTIALS

The monograph is organized into three parts. The first part entails Chapters 2 to 5 and contains information that I consider essential both for forming an opinion about results obtained by others, as well as for the planning and analysis of an observational Before-After study. All chapters of this part have been written without any mathematics and are therefore accessible to all.

Chapter 2 discusses the basic logic behind all attempts to estimate the effect of some treatment. It shows why habits well suited for the physics laboratory do not fit the reality of research on road safety. In the third chapter I attempt to define what is to be meant by 'safety' and how it differs from 'security'. I then proceed to argue that in evaluative research on safety, accident frequency is to be preferred to accident rate. Chapter 4 describes the properties of the raw material - the accident counts: what a motor vehicle accident is, what makes it reportable, how many are reported, and how much of the recorded information is correct. I also discuss the elusive concept of 'target accidents'. Chapter 5 deals with the role of prediction and estimation. To find what effect some treatment had on safety, one must predict what safety would have been without the treatment and compare this to an estimate of what safety was with the treatment in place.

CHAPTER 2
WHAT IS THE QUESTION?

The structure of a Before-After study is deceptively simple. To determine the effect of heat on the length of a copper rod in a physics laboratory, we measure the length of the rod, heat it, and measure again. The difference between the two measured lengths is due to the heat. Since all measurements are subject to some error, the effect of heat on the length of a copper rod is not known exactly; it is an estimate. Much of our intuition is rooted in this kind of experiment. The idea behind a Before-After study in road safety is very similar. To determine the effect of some treatment we measure safety before the treatment by counting the number accidents in the 'before' period and then measure again by counting the number of accidents in the 'after' period. If nothing else changed, the difference is attributed to the treatment. The count of accidents as a measure of safety is a measurement just as length is, and it is also subject to error. Thus the change in safety which is due to the treatment is also not known with certainty.

To illustrate, consider the results of the R.I.D.E.[1] enforcement program to reduce impaired driving introduced in one of five police districts of Metropolitan Toronto as shown in Figure 2.1.

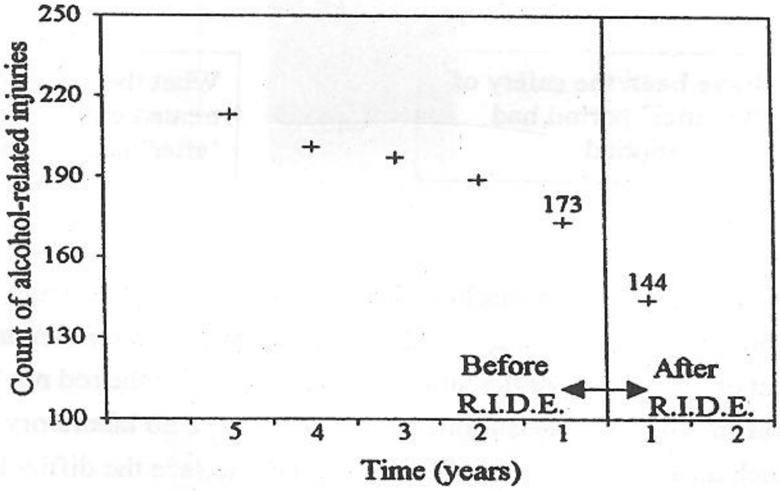


Figure 2.1. Time series of accidents counts. The R.I.D.E program.

---

1 Reduce Impaired Driving Everywhere.

One year before the R.I.D.E. program started, 173 persons were injured in alcohol-related accidents. During the first year with the R.I.D.E. program in place, 144 persons were injured in such accidents (Vingilis et al., 1979). The reduction was of 29 injuries (with a standard deviation of about 18). Can the difference, however uncertainly known, be attributed to R.I.D.E.? While in the physics laboratory this could be done, to do so in road safety, I will claim, would be wrong.

The reason for which one cannot transplant the logic of the physics laboratory into road safety becomes apparent upon the inspection of Figure 2.1 in which the plus signs represent a time series of accident counts. Naturally one asks: what would have been the number of alcohol-related injuries in the 'after' year, had R.I.D.E. not been in place? It does not seem proper to guess that it would be the same as in the 'before' year because there is an apparent time trend. In contrast, in the physics laboratory it is entirely reasonable to assume that had the copper rod remained unheated, its length would not have changed. We could easily convince ourselves of this by repeating the 'before heating' measurement on several occasions. If we do not bother doing so, it is because in our experience the passage of time has little to do with the length of the copper rods. Yet in road safety this kind of assumption is obviously false. The safety of everything changes with time. One may not assume that had no treatment been applied, safety in the 'after' period would have been the same as in the 'before' period. Therefore, to assess the effect of a treatment on the safety of some entity one has to compare what would have been the safety of the entity in the after period had treatment not been applied, to what the safety of the treated entity in the after period was.

Thus, the logical basis of any inquiry about the effect of any treatment is in the comparison of:

What **would have been** the safety of the entity in the 'after' period had treatment not been applied	with	What the safety of the treated entity in the 'after' period **was**.

This logical basis applies to physics just as it applies to road safety. Only in physics, because one works in the laboratory where all other conditions are supposed to remain unchanged, one may often take the shortcut of assuming that the 'after' measurement, had the rod not been heated, would have been the same as the 'before' measurement. We, who have no laboratory, cannot make such shortcuts or make such an assumption. In all cases we have to face the difficult task of predicting 'what would have been'. Without doing so there is no valid comparison; without a valid comparison, there is no estimation of safety effect. For further discussion see the endnote.

It is now easy to see why many are critical of simply comparing the count of accidents 'before' to the count of accidents 'after' and then attributing the difference to the treatment. The tacit assumption of such a naive comparison is that the count of 'before' accidents is a good estimate of what would have been the count of 'after' accidents. But since traffic, weather, road user demography, car fleet and other important factors all change in time, the tacit assumption is obviously incorrect. Therefore, the simple comparison of before and after accident counts reflects not only the effect of the treatment but also the effect of changes in all the other factors.

By specifying what is the basic logic of inquiry I have also given meaning to the phrase: 'the effect of some treatment on the safety of an entity'. The 'safety effect of a treatment' is the difference between what safety would have been in the absence of treatment, and what safety was with the treatment in place. This can be shown as in Figure 2.2.

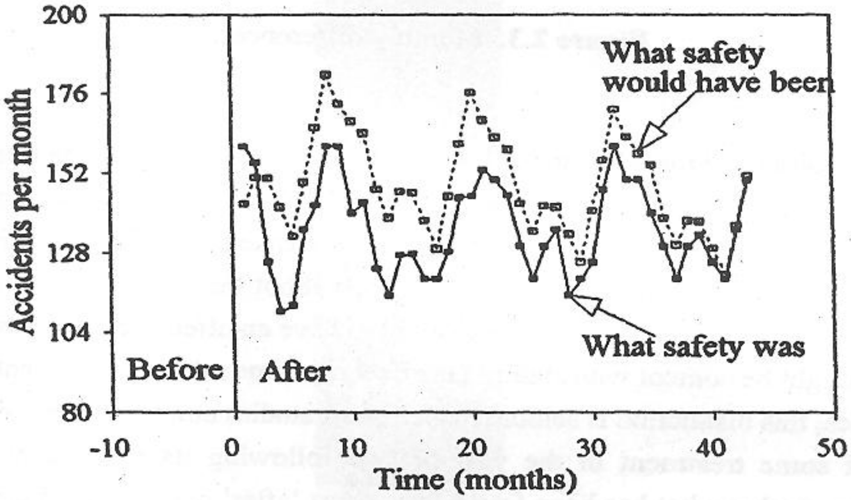


Figure 2.2. How the safety effect varies over time.

The treatment was implemented at time zero. What safety would have been in the subsequent months had the treatment not been implemented is shown by empty squares (joined by the dashed line); the full circles (joined by the solid line) represent what safety was with the treatment in place. In the case shown, the effect of the treatment is beneficial except during the first two months after implementation - the adjustment period. The beneficial effect is seen to fade gradually. In Figure 2.3 I show the differences between the prediction of what would have been and the estimate of what was. The total safety effect of the treatment would be the sum of the monthly differences.

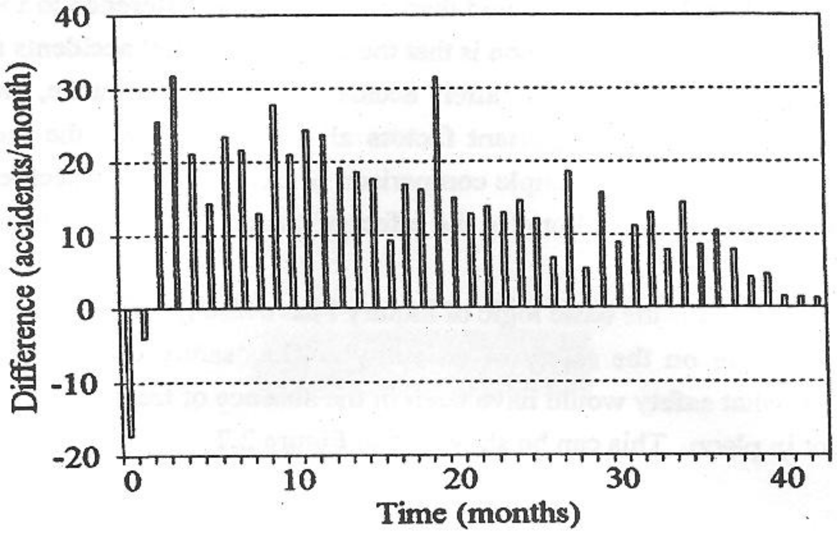


Figure 2.3. Monthly differences.

The main point of Figures 2.2 and 2.3 is to show that the effect of some treatment on safety is, in principle, not a single number. For some treatments (e.g., pavement resurfacing) it is natural to think that the effect on safety fades with time. In such a case one must speak about the effect in months (years) 1, 2, 3, . . . after treatment, or perhaps about the sum of these. Other treatments (e.g., replacing stop signs by signals) may be thought to have an effect that persists indefinitely. In such a case one might be content with stating the effect as a time rate; e.g., accidents saved per year. In present practice, this distinction is seldom made. Most studies aim to estimate what has been the safety effect of some treatment in the year or two following its implementation. The tacit assumption seems to that what has been found for a short 'after' period would persist indefinitely. This is a questionable assumption. First, there is road user adaptation, a subject that is at present ill understood and full of surprises. Second, there are changes in time to the treatment itself (e.g., gradual polishing of the crushed stone aggregate after resurfacing). For these reasons it is prudent to remember that the safety effect of any treatment may be changing with the passage of time.

So far I spoke about the 'safety of an entity' and the 'safety effect of treatment' without declaring what safety is to mean and how it can be measured. I have said that it is to be measured by the count of accidents but did not elaborate. The task of defining safety can no longer be postponed; it is the subject of the next chapter.

Suggested Readings.

1. About 'validity' and threats to it, pages 1-12, in Campbell, D.T., and Stanley, J.C., (1966). *Experimental and quasi-experimental designs for research.* Rand McNally College Publishing Company, Chicago. A later and expanded version is Cook, T.D. and Campbell, D.T. (1979). *Quasi-experimentation, Design and analysis issues for field settings.* Rand McNally, Chicago and Houghton Mifflin, Boston. Bias in observational studies is discussed in Rosenbaum, P.R., (1995). *Observational studies.* Springer Verlag, New York.

2. About 'validity', 'causality' and the 'scientific method', see Chapters 1 and 2 in Kleinbaum, D.G., Kupper, L.L., and Morgenstern, H. (1982). *Epidemiologic Research*, Lifetime Learning Publications, Wodsworth Inc., California.

3. About causality and comments on the debate about the link between cancer and smoking, see Section 9.5 in Spirtes, P., Glymour, C., and Scheines, R. (1993). *Causation, Prediction and Search.* Springer-Verlag Inc., New York.

4. Surgeon General of the United States (1964). *Smoking and Health.* U.S. Government Printing Office.

Endnote.

Suppose that a study found that during the first year after the pavement of a set of roads has been resurfaced, there were x% more accidents than what would be expected in the same period had they not been resurfaced. Can one say that the resurfacing was the 'cause' of the increase, and can one believe that if the pavement of another set of such roads was resurfaced a similar increase is to be expected? The issue requires airing.

The claim has been made (in the text) that the effect of a treatment is revealed by comparing what would have been the safety in the after period had the treatment not been implemented, with what safety in the after period was. Hidden in this claim is an assertion about cause and effect, namely: if there is an effect there must be a cause and, if all remains constant except one 'thing' and there is an effect, then that 'thing' is its cause. At the semantic level this assertion is quite unobjectionable to most. However, the 'unobjectionability' of the assertion is bought at the expense of an impossibility. It is quite impossible to know 'what would have been' for reasons that are both practical and conceptual. Even those who think that the world is built along cause-and-effect lines will concede that any attempt to predict what would have been must rely on an incomplete understanding and description of a complicated, perhaps endless, causal web. This is why, in spite of the seductive simplicity of the claim and the appearance that, as a consequence, causal inferences are unobjectionable, a word of caution is in order.

The culture in which physical scientists and engineers grow up does not emphasize concern about cause and effect. In endeavors in which much work is done under controlled conditions and repeatability of results is the norm, there is little reason to worry about what seems to be unnecessary metaphysics. Mathematicians and statisticians also have little reason to dwell on such issues. Their work is conducted in the universe of deduction, not the real world. In contrast, the disciplines such as social science, psychology, economics, etc. worry about definitions and meaning. In these disciplines complexity of interaction and uncontrollability of circumstance are the rule. Therefore, a great deal of effort goes into polishing terms and concepts in the hope that such care will help to discover some general truth, making for reliable prediction of effects from causes. Our endeavor, the aim of which is to learn about the safety effect of treatments, comes close to the circumstance of the social scientist. Thus, concern about when one can make inferences about cause from estimates of effect should be foremost in our mind. However, our intellectual heritage (speaking as an engineer), does not tell us to raise this question.

Returning now to the resurfacing example, we had to predict what would have been the number of accidents on these roads in the year after resurfacing had resurfacing not been done. The question to ask is: has the study correctly accounted for all the possible (or even just important) influences? For, if not, one would be incorrectly attributing to resurfacing the effect of such unaccounted for factors. In any real study the requirement to **know** what all important influences are, and to have their **measure**, and also to **account** for their effect correctly is an unattainable goal. Therefore one cannot be sure that studies of this kind identify cause.

For these and similar reasons, other disciplines (e.g., epidemiology) provide guidance on when causal judgement may be justified. Thus, when the Surgeon General (1964) asserted that smoking causes cancer, the following ad hoc rules for judging causality were used:

1. Strength of Association (meaning some statistical measure of association is strong).
2. Dose-response effect. (The more of the causal factor, the larger the effect.)
3. No temporal ambiguity. (Disease follows exposure to risk factor).
4. Consistency of findings. (Several studies produce similar results).
5. Biological plausibility. (The hypothesis makes sense in view of what is known in biology).
6. Coherence of evidence. (Some combination of 4 and 5).
7. Specificity. (Causal factor causes this disease, and this disease is due to this causal factor).

Many of these rules are deficient and some have no bearing on issues in road safety. However, some adaptation of 1, 2 (where applicable), and of 4 and 5 would be both timely and desirable.

CHAPTER 3
DEFINING SAFETY

The goal is to estimate what effect a treatment had on safety. Naturally, I have to say what 'safety' is to mean and how it can be measured. This may seem as unnecessary pedantry; it is known in advance that eventually I will speak about the number of accidents, number of injuries and of their severity. It turns out, however, that while safety and the count of accidents are related, they are not synonymous. If the difference is not clearly stated, misunderstanding, confusion and unnecessary controversy will abound. However, even before the meaning of safety can be defined, some underbrush has to be cleared.

3.1 UNDERBRUSH

The principal manifestations of road safety (perhaps 'road unsafety') are accidents and their harm. However, road safety may also be taken to mean something like the feeling of personal safety when on the road. In fact, it is possible that some treatment induces in the road user a false sense of safety and, as a result, the number of accidents may increase. Thus, while the road user thinks that the road has been made 'safer', the treatment has made it less 'safe'. Clearly different words are needed for these two different notions:

a. the objective measure reflected in the prevalence of accidents and their harm and,
b. the subjective perception of how safe one is on the road.

I will stick with the convention that **road safety** is manifest in the occurrence of **accidents** and their harm; I will refer to people's **subjective perception** of safety as the feeling of **security.**

To illustrate, let the abscissa of A in Figure 3.1 describe the pedestrian safety if crosswalks at intersections are not marked by painted lines, and let the ordinate of A measure the pedestrians' feeling of security under such conditions. Let A' represent the state of affairs after crosswalk edge lines have been painted. Pedestrians now feel protected by the two lines of paint on the pavement and therefore their security has increased. However, their risk of being run over may have increased (at least in accord with what has been found in San Diego by Herms, 1972). Pedestrians may have been lulled into a false sense of security.

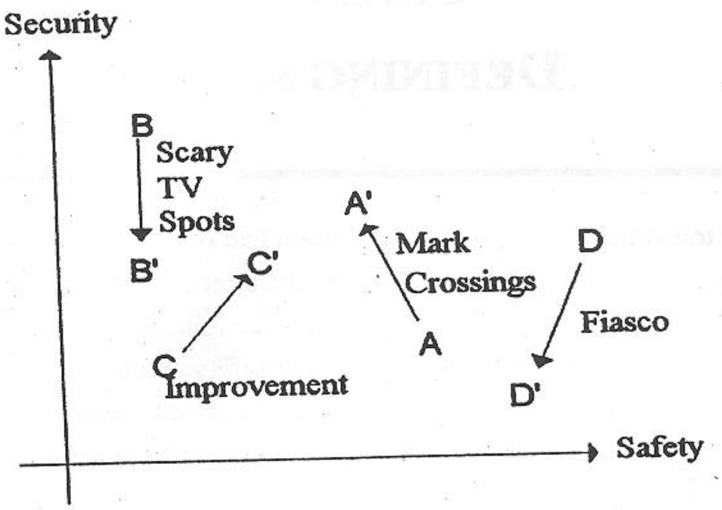


Figure 3.1. Changes in 'safety' and 'security'.

In the same figure, consider a publicity campaign that consists of showing on TV the mangled bodies of accident victims, placing demolished cars in the center of roundabouts, and posting the number of injured persons alongside many road sections. This kind of campaign is not uncommon; I have witnessed one lately. Let B represent the state of affairs before the campaign and B' be the safety and security after the campaign has run its course. Road users now feel less safe and the number of accidents may have dropped very slightly. One could add a story in which B' is shifted to the right indicating improved safety. A local road constriction common in traffic calming may have such an effect. Neither the shift from A to A' nor B to B' is a clear-cut improvement. The A→A' treatment made people feel more secure but increased accidents; the B→B' treatment made people more anxious about their personal security while the number of accidents may have decreased slightly. A clear-cut improvement is if the change is of the C to C' kind. A clear-cut fiasco is a change from D to D'. Real-life treatments may be of the A, B, C or D kind. Therefore, both changes in safety and changes in security should be measured. In this monograph, only the question of changes in safety will be discussed.

The next question is whether the concept of safety must be linked to accidents. I relied only on common convention when I claimed that accidents (crashes) and their harm are the principal manifestation of road safety. Safety might be given an alternative meaning by recognizing that each accident is preceded by a dangerous situation; some turn into accidents, the rest into near-misses. Each dangerous situation, in turn, is preceded by some incipient danger. In fact, one can picture the continuum of events preceding accidents as the pyramid in Figure 3.2.

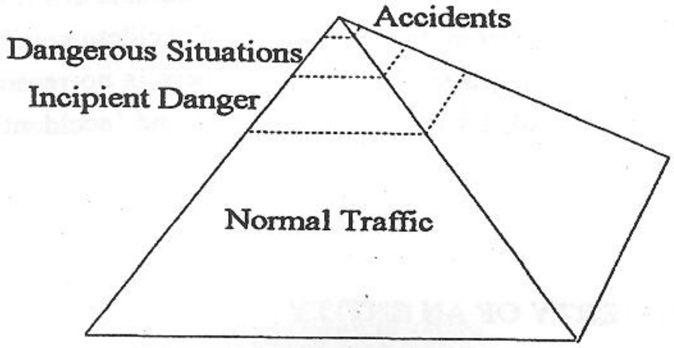


Figure 3.2. The continuum of events.

The volume of each layer represents the frequency of the events. Thus, there are few accident events per unit of time but many more dangerous situations or events. Since what occurs more frequently is easier to measure, there is a great temptation to define safety in terms of the frequency of dangerous events. This, in fact, was the idea behind the use of so called 'surrogate measures of safety'. One of those has been converted into an operational procedure - the Traffic Conflicts Method. By observing change in the frequency of conflict occurrence (e.g., before and after some treatment) one hopes to be able to determine changes in safety.

The hope rests on the observation that where there is smoke there is fire. The usual assumption is that the frequency of dangerous situations (conflicts) increases in proportion to the frequency of accidents. If the assumption is true, then dangerous situations (conflicts) can measure safety because they are proportional to the mean accident frequency. In this case, safety is again linked to accidents. If this assumption is not true, if it were possible to have an increase in the mean frequency of dangerous situations without an increase in the mean accident frequency, one would not think that safety can be measured by the frequency of dangerous situations. It follows that, in the final account, to preserve the ordinary meaning of words, the concept of safety must be linked to accidents.

Another semantic issue is the use of the word 'accident'. Those who prefer to use 'crash' instead, think that the word 'accident' has connotations of it being an unavoidable event[1]. This, they fear, might weaken the resolve to intervene in order to reduce crashes and their harm. But our context is that of measuring the safety effect of treatments. In such a context a fatalistic interpretation of 'accident' can make no sense; why intervene if accidents are preordained or unavoidable? Thus, there is no danger of mis-communication here. One good reason for not

---

[1] Evans argues that: "The word *crash* indicates in a simple factual way what is observed, while *accident* seems to suggest in addition a general explanation of why it occurred. . .."

abandoning the term 'accident' is that it is the common currency in the community of transportation engineers to whom this monograph is directed. Another good reason is that much of the discussion will be of the statistical issues, arising from the randomness of accident counts. For this purpose, the word accident provides the proper associations. But, there is no reason to be diverted by doctrinaire polemics. To placate all, I will use both 'crash' and 'accident', and treat them as synonyms.

3.2 SAFETY AS A PROPERTY OF AN ENTITY

What then should we call the safety of a certain intersection, driver or city? Listed in Table 3.1 are the counts of police-reported accidents at a Toronto intersection during 52 Police Reporting Periods. A year is divided into 13 Police Reporting Periods (PRPs) each being 28 days in duration.

Table 3.1. Reported accidents at the Eglinton Ave. and Don Mills Rd. intersection. January 1, 1970 - December 31, 1973

1970		1971		1972		1973	
PRP	# Acc.	PRP	# Acc.	PRP	# Acc.	PRP	# Acc.
1	5	14	2	27	3	40	5
2	3	15	3	28	5	41	5
3	1	16	4	29	8	42	5
4	4	17	1	30	4	43	3
5	0	18	3	31	1	44	6
6	3	19	2	32	2	45	1
7	5	20	1	33	6	46	6
8	1	21	4	34	4	47	2
9	3	22	2	35	6	48	2
10	6	23	5	36	8	49	3
11	1	24	0	37	4	50	5
12	8	25	5	38	9	51	6
13	0	26	3	39	6	52	9

These accident counts are shown in Figure 3.3 and are seen to jump all over the place. Thus, the most important impression about the count of accidents is their unsteadiness.

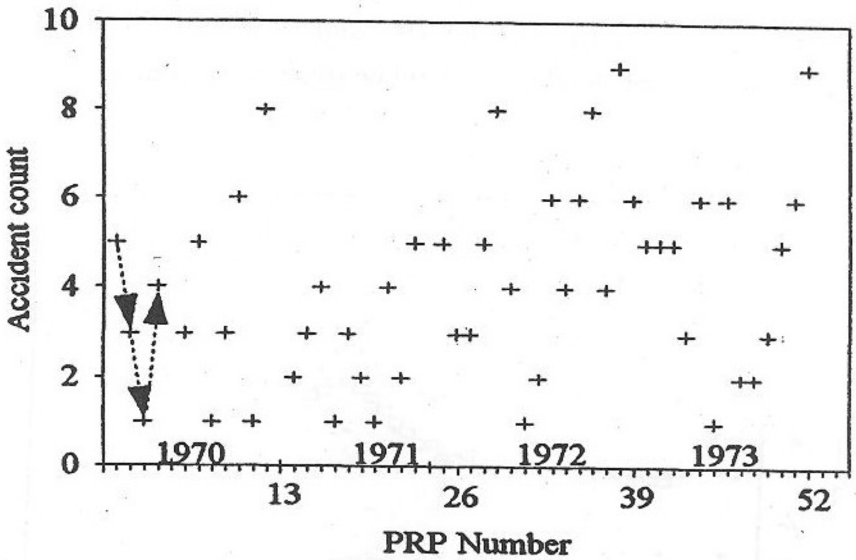


Figure 3.3. Count of accidents in police reporting periods.

Obviously, one cannot equate 'safety' with the 'count of accidents'. Was one to do so, it would mean that safety improved from PRP 1 to 2, and also from PRP 2 to 3 because the count of accidents dropped from 5 to 3 to 1; but then safety deteriorated in PRP 4, improved again in PRP 5 and so on. Taking this ad absurdum, days with accidents would be considered unsafe, days without accidents - totally safe. It is obvious that equating safety with the vagaries of accident counts is unhelpful. What we see are random fluctuations of accident counts that cannot be attributed to causes of interest, nor can they be linked to 'treatments' such as channelizing the intersection or changing the duration of the amber phase in its signal. I need to coin a concept of safety so that it is a fairly stable property of the entity under scrutiny. It needs to be coined so that if the entity does not change and if its environment, users, and level of use remain the same, then its safety ought to remain unchanged. To use a common analogy, when a die is thrown several times and every time a different face shows, we do not think that between each throw there has been a change in the nature of the die. The stable property of a fair die is the fact that each face has an equal chance of showing up. This stable and characteristic property of a die can be ascertained only by repeated throws. For a fair die, as the number of throws increases, the proportion of time a certain face shows approaches 1/6 and the long-term average score is 3.5. Thus, the stable property of a die is revealed only in a long sequence of throws. What might be a corresponding characteristic and stable safety property for the intersection with accident counts in Table 3.1 and Figure 3.3?

In Figure 3.4, the number of accidents recorded in a police reporting period is shown as a plus sign. One cannot see much rhyme and reason in this cloud. Let us, however, start calculating running averages. In Figure 3.4 the average for the first 13 PRPs is shown by a square above PRP 7, the average for PRPs 2 to 14 is plotted above PRP 8 and so on. Note that these averages seem to be more stable. Thus, behind the cloud there might be some permanent and stable property after all.

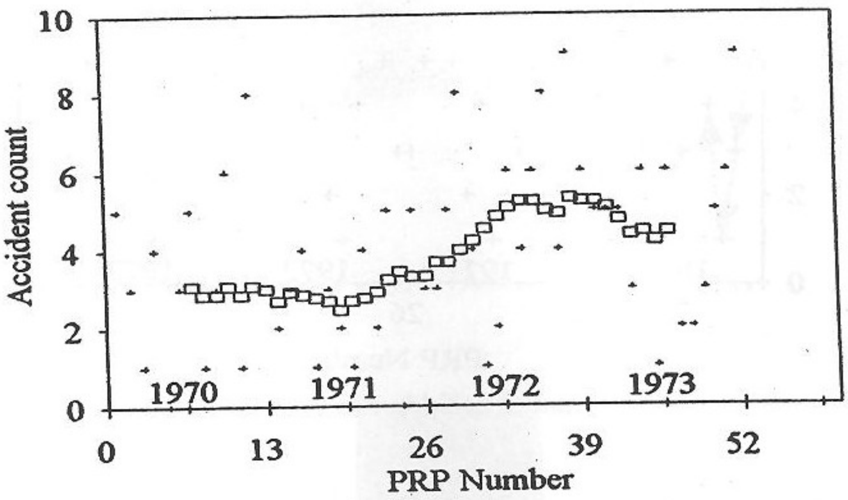


Figure 3.4. Running averages of accident counts.

As for the die, we choose the characteristic property to be some 'average in the long run'. From an inspection of the squares in Figure 3.4 it is clear that, however safety will be defined, it must be allowed to change over time. Therefore, I will have to carefully describe what 'average in the long run' is to mean. Even though safety has not yet been clearly defined, an important distinction can already be made. The distinction is between the real accident counts (the plus signs) which are directly observable, and between the long-term-average lurking in the background, that cannot be observed directly. It is this ephemeral long-term-average that will represent the stable safety property. The squares in figure 3.4 are hints about what these invisible long-term averages might be.

To illustrate the importance of the distinction between the count of accidents and safety, consider the following quote from the Toronto Globe and Mail newspaper in which a cabinet minister was quick to take credit for the effect of a law that he championed:

> "A new law allowing Ontario police officers to suspend temporarily the licenses of marginally impaired drivers contributed to a 23 per cent reduction in traffic deaths . . ., statistics show 47 people were killed in traffic accidents in January (1982), down from 61 in the same month last year."

It is quite possible that the new law did improve road safety in Ontario. However, whether this is true becomes moot when one examines Figure 3.5. Figure 3.5 shows the count of fatalities[1] in a time series of Januaries. These counts are subject to the same kind of randomness as that shown in Figures 3.3 and 3.4. The claim in the quote is based on the reduction shown in the ellipse. Such drops between one January and the next are not very unusual; note, e.g., the arrows for 73/74 or 75/76. Similarly, one sees fairly large increases when the arrow points upward (except that there is no rush to take credit for these). The moral is this. We deal with a phenomenon that is subject to randomness. That is, even when the main properties of the entity did not change, there are always up and down jumps in accident counts. Therefore, the attribution of a specific jump that is similar to all the other unexplained jumps to some specific action or event (here the new law in the quote) is inappropriate.

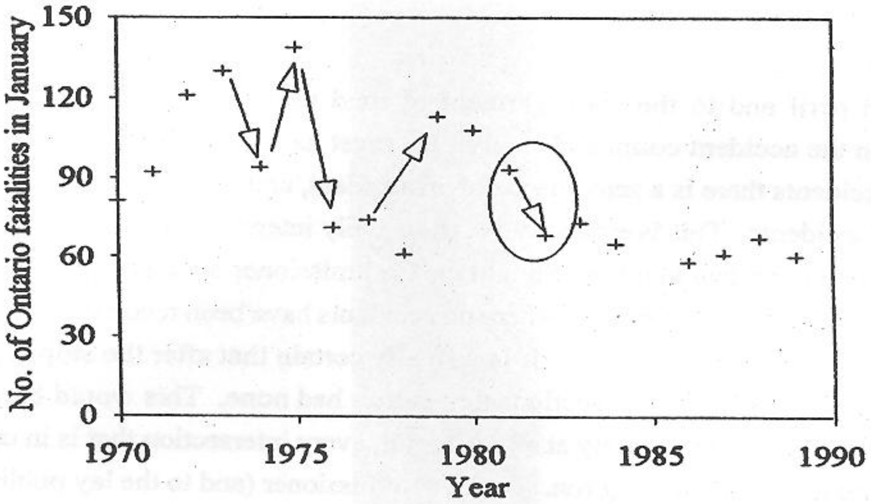


Figure 3.5. Ontario fatalities in a time-series of Januaries.

Those who advised the minister failed to distinguish random fluctuations in accident counts from changes in the underlying but invisible property of safety. The ups and downs in the plus signs of Figure 3.5 have no more meaning than the corresponding fluctuations in Figure 3.4. Any opinion about change in safety must refer to the 'squares' - the underlying 'long-term-average'. Was the 1982 count of fatalities a hint that in that year there was a downward shift in the underlying but unknown time series of squares? What was the likely size of such a shift and how

[1] There is a large discrepancy between the numbers in the quote and those in Figure 3.5 which were taken from official statistics published later. This serves to underline the fact that even in a well organized jurisdiction the count of fatalities is often uncertain.

accurately can its size be estimated? These are the relevant questions. Obviously, they cannot be answered simply[1].

Another example of the confusion and harm of failing to distinguish between the count of accidents and safety came across my desk in the form of a memorandum by the Commissioner of Public Works to the City Services Committee (Commissioner, 1991). As so many traffic engineers are wont to do, he pushed for the removal of 'all-way' stop signs. His staff told him that research shows that when 'two-way stop' control was replaced by 'all-way stop' control, the average number of accidents was halved. Therefore, a change back to two-way stop control may harm safety. But, the Commissioner noted that: "accidents may only be reduced (by using all-way stops) at locations where, in fact, accidents have occurred." By implication, safety cannot be harmed if an intersection has had no recent history of accidents. At such sites, so the Commissioner thinks, the removal of all-way stops cannot do harm.

To his own peril and to the endangerment of road users, the Commissioner failed to distinguish between the accident count and safety. He must have thought that at an intersection with no recorded accidents there is a zero chance of an accident, and, as a consequence, also a zero average number of accidents. This is evidently not true. Only intersections not open to traffic can have a zero chance of an accident to occur. Should the Commissioner's faulty logic eventually lead to the removal of stop signs at intersections where no accidents have been recorded before, he could have a future embarrassment on his hands. It is virtually certain that after the stop sign removal, there will be some accidents at the intersections that earlier had none. This would happen even if the stop sign removal did not affect safety at all. After all, every intersection that is in use has some non-zero accident frequency in the long run. To the commissioner (and to the lay public) who tend to equate safety with count of accidents, the occurrence of some accidents where there were none before would look like a degradation in safety.

In short, safety is not to be equated with the fluctuating accident counts; rather, we should define safety as an underlying stable property that has the nature of a long-term average. One way to define the safety of an entity is as:

> *the number of accidents (crashes) by kind and severity, expected to occur on the entity during a specified period.*

---

[1] A procedure for doing so is suggested in E. Hauer (1996). Detection of safety deterioration in a series of accident counts. Transportation Research Record 1542. Transportation Research Board. Washington.

For clarity in communication, it may be useful to define safety not as 'number of accidents', but as 'number per unit of time' (for example, accidents/year). In this standardized form, **safety is an expected accident frequency**. This nonessential modification of the definition consists of merely dividing the expected number by the duration of the 'specified period'. Both the expected number and the expected frequency versions will be used.

The term 'expected' is used here as in the theory of probability and corresponds roughly to 'average in the long run'. However, as noted earlier, the meaning of the 'long run' requires further specification. It is essential to allow for the safety of an entity to change over time. Therefore 'averaging over the long run' cannot mean the averaging over many consecutive periods. To complete the definition of safety, the phrase 'long run' has to be given meaning.

The squares in Figure 3.4 give a strong indication that the safety of that intersection was changing with time. Since safety changes with time, there can be no real 'long-run' in the usual sense. Accordingly, one has to interpret the term 'expected' in the definition as: 'what the limit of the long-term average would be, if it were possible to freeze all the relevant conditions of the specified time-period, and repeat it many times'. By saying: ". . . would be . . ." and by adding the condition ". . . if it were possible . . . ," I have elected a definition of safety that is logically defensible but empirically troublesome. Neither can conditions be frozen nor can periods be repeated. Therefore, only one observation (accident count) is possible in each period. This makes the statistical estimation difficult. Unfortunately, I do not know of a better definition. C'est la vie.

Sometimes statements about safety are made in terms of accident frequencies (e.g., fatal accidents/year), sometimes in terms of consequence frequencies (e.g., fatalities per year). Since both kinds of statements are useful a broader definition of safety is:

> *the number of accidents (crashes), or accident consequences, by kind and severity,* ***expected*** *to occur on the entity during a specified period.*

In summary, the safety of an entity is a series of **expected** numbers or frequencies[1], one for each accident type or crash severity. These expected numbers change in time. Safety is not to be equated with the count of accidents. Counts are always non-negative integers; expected numbers can be any non-negative numbers, integers or not. The count of accidents is but a reflection of the underlying expected number, a hint that enables us to estimate what the expected number or frequency at some point in time is or was.

---

[1] It is a frequency when we speak of accidents/unit of time.

3.3 FREQUENCY OR RATE?

I have defined the safety of an entity to be its set of expected **numbers or frequency** of accidents of specified type or severity prevailing in a specified period. For a certain accident type and severity,

$$E\{Accident\ Frequency\} \doteq E\{Number\ of\ Accidents\ on\ entity/Unit\ of\ time\}$$

The notation E{.} stands for 'expected value'. The entity may be a road, a road section 1 km long, an intersection and so on.

Current engineering practice is to measure safety not by expected accident frequency, but by the expected **accident rate.** Accident rate is defined as:

$$Accident\ Rate \doteq \frac{Accident\ Frequency}{(Exposure/Unit\ of\ time)}$$

Exposure is usually derived directly from traffic flow. Thus, e.g., if exposure is to be defined as vehicle kilometers of travel in a year, you multiply the annual average daily traffic (AADT) by 365 (or 366) and by the length of the road section. To illustrate, if the accident frequency on a 2 km road section is 5 accidents per year and the AADT is 10,000 vehicles per day (vpd), the accident rate is $5/(10000\times365\times2)$=0.68 accidents/million vehicle-kilometers. The question is:

> *What measures the safety effect of a treatment? Is it the change in expected accident frequency or the change in the expected accident rate?*

The main aspects of this question were examined by Pfundt (1969), Hakkert (1976), Mahalel (1986), Brundell-Freij & Ekman (1991), Hauer (1993, 1995 b) and perhaps by others. Empirical evidence shows that the relationship between expected accident frequency and traffic flow is usually not a linear one. Logical considerations also lead to that conclusion. To illustrate, if the number of parked-on-shoulder vehicles is proportional to traffic flow, and the number of vehicles passing the parked-on-shoulder vehicles per unit of time is the traffic flow, then, at least for low traffic flows one may expect the number of collisions with parked-on-shoulder vehicles per unit of time to be proportional to the square of traffic flow. Similarly, at low flows one may expect the number of single vehicle accidents to be proportional to flow. But, as traffic flow increases, it becomes more and more difficult not to hit another car. Therefore, for single vehicle accidents, proportionality cannot be expected to hold at high flows. Assume then that the relationship between the expected frequency of, say, injury accidents and traffic flow for two-lane rural roads is curvilinear as shown in Figure 3.6. This reflects mainly the fact that drivers behave differently in sparse and heavy traffic and that frequency of single and multi-vehicle accidents depends on traffic flow, speed and density.

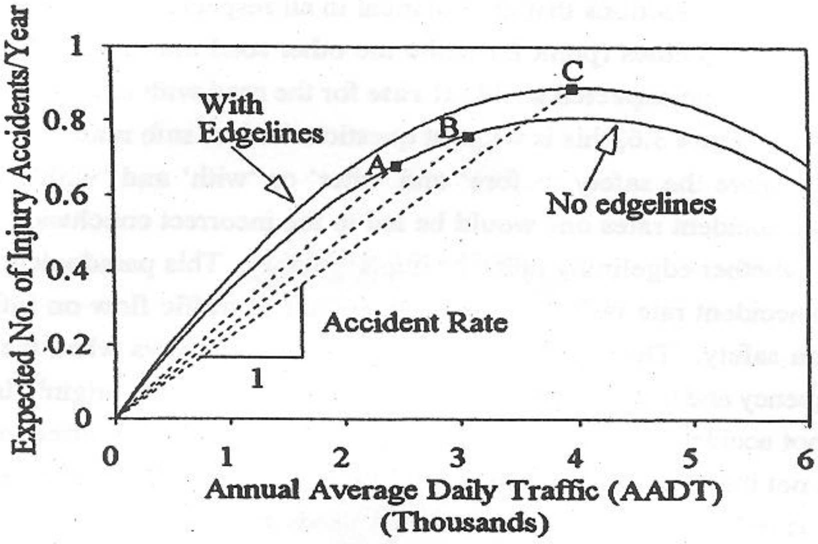


Figure 3.6. Relationship between traffic flow and safety.

Let the ordinate of point A represent the expected number of injury accidents per year on a rural two-lane road without edgelines at AADT=2500 vpd. Thus, 0.69 injury accidents/year is the expected **accident frequency** for this traffic flow. Note that in this representation, the expected **accident rate** for AADT=2500 vpd is proportional to the slope of the line joining the origin and point A. Imagine now that during a certain 'after' period, the road and all other conditions remained the same as during the 'before' period, but the AADT increased to 3000 vpd. This corresponds to point B. The slope of the line joining B to the origin is smaller than the slope of the line to point A; that is, the expected **accident rate** has diminished. Does it mean that the road is now safer?

From an individual motorist's point of view, travel is safer because the chance of being in an injury accident has diminished. However, the highway engineer cannot claim that the road has been made safer, for nothing was done to it. Since the accident rate is seen to be changing even without any treatment, it cannot serve to measure the safety effect of treatments. In particular, it is incorrect to believe that use of the accident rate somehow accounts for difference in traffic flow before and after treatments. It would do so only in the special case when it can be assumed that the expected accident frequency is proportional to traffic flow.

The danger of using the accident rate for describing the effect of treatments can be better illustrated by comparing two roads that are identical in all respects, except for edgelining. As shown in Figure 3.6, the road with the edgelines is expected to have more injury accidents at any traffic flow than an identical road without edgelines; it is without question less safe[a].

Consider now two road sections that are identical in all respects except that one road has an AADT of 3000 and no edgelines (point B) while the other road has an AADT of 4000 and painted edgelines (point C). The **expected accident rate** for the road with edgelines is seen to be smaller, yet, as is clear in Figure 3.6, this is without question the less safe road. Thus, the use of the accident rate to compare the safety 'before' and 'after' or 'with' and 'without', leads to a paradox; by comparing accident rates one would be led to the incorrect conclusion about which road is safer or about whether edgelining helps or impairs safety. This paradoxical result stems from the fact that the accident rate fails to separate the effect of traffic flow on safety from the effect of edgelining on safety. This kind of error will be present always when the relationship between accident frequency and traffic flow is not a straight line through the origin[b]. In conclusion, accident frequency, not accident rate, should be used to measure the safety effect of treatments. This, of course, does not imply that changes in traffic flow from 'before' to 'after' should not be taken into account. It only implies that such change needs to be accounted for correctly, by modifying the accident frequency. How this is to be done will be discussed in Chapter 8.

3.4 CHAPTER SUMMARY

If one is to measure how safety has changed due to an intervention, it must be clear what safety is. In lay circles it is common to equate safety with the count of accidents or with the derived measures of accident frequency and accident rate. For our purpose, it is impractical to do so. The count of accidents changes from one period to another even when there has been no change in any observable causal factor. It is the essential nature of accident counts that they are subject to randomness. It follows, that the only useful definition of safety is in terms of the elusive 'mean' or 'average in the long run' that is behind the randomly fluctuating counts. Accordingly, I define the safety property of an entity as the number of accidents, or the number of accident consequences, by kind and severity, expected to occur on the entity during a specified period.

Some additional verbal fencing is needed to lessen the chance of miscommunication. First I draw a distinction between the safety property of entities and the perception of safety by road users, which I call security. Second, I point out that events preceding accidents, traffic conflicts, dangerous situations and the like, can not measure safety unless they can be shown to be related to safety as defined above. Third, I offer my reasons for sticking by the word 'accident' instead of the now politically correct 'crash'.

Another contentious issue discussed is whether safety should be defined as expected frequency (count/unit of time) or as expected rate (count/unit of exposure). This is a difficult argument to make because the use of accident rates is a time-honored habit and, to some, an article of faith. I show that when safety is clearly improved the expected accident rate may increase. This fault disqualifies expected accident rate as a legitimate measure of safety. in contrast, expected accident frequency is free of such ambiguity. That expected accident frequency is to be used does not mean that differences in exposure can be neglected, only that they have to be accounted for correctly.

Endnotes.

Endnote a. The painting of edgelines is often thought to enhance safety. However, the studies on which this belief is based are often seriously flawed. One of the best studies concludes that: "There is still no evidence that edgelines had reduced accident frequencies. In fact, there are indications that the sites with mixed pattern (broken and continuous lines) experienced an increase in total (injury) accidents ... relative to control (sections), but the evidence for this is not particularly strong" (See News, 1985, p.85). The painting of edgelines does increase 'security'. Perhaps it is like going from A to A' in Figure 3.1. It may still be deemed worthwhile, because it is important to make road users feel safer. However, it is unclear where is the right balance when making people feel more secure while increasing their chance of accident occurrence.

Endnote b. Many highway and traffic engineering standards are tied to a 'volume warrant'; for roads below the volume warrant a lower standard is used. Imagine that there is a 'volume warrant' for edgelining and a research study is commissioned to compare accident rates of roads with and without edgelines. Because of the existing volume warrant, high volume roads are equipped with edgelines. If the performance curves are curvilinear, as in Figure 3.6, the higher the volume the lesser the accident rate. Thus, the study will find that roads with edgelines have a smaller accident rate. This would be so not because edgelines add to safety but because higher volume roads have a smaller accident rate and these tend to be equipped with edgelines.

The danger is clear. The use of accident rate as a measure of safety when coupled with the widespread use of volume warrants may lead systematically to erroneous conclusions. This is another brand of 'bias by selection'. The term has been originally coined to represent the bias which arises when entities with an unusually high number of accidents are selected for treatment and this same number of accidents then is used in the 'before' period. Here we witness a similar bias which is due to selection by a volume warrant. Fortunately, this latter bias can be avoided by shunning the use of accident rates and using accident frequencies instead.

CHAPTER 4
COUNTING ACCIDENTS

Safety has been defined to be an expected accident frequency. Expected values, being akin to an 'average-in-the-long-run' are never known, they can only be estimated. The count of accidents (crashes) is the raw material from which estimates of safety are produced. Naturally, one has to know the raw material to know the quality of the product. Two important questions arise:

1. *What accidents are being counted and are therefore available for safety estimation?* What counts as a motor vehicle accident is a matter of definition. Not all motor vehicle accidents are reportable, not all reportable accidents are reported, and not all reported accidents are correctly recorded.

2. *Which accidents should be considered in a Before-After study?* If the treatment is, say, that of illuminating a stretch of road, noontime accidents can be hardly affected; or can they? Only some accidents are the target of a treatment. But to answer the question is not simple.

4.1 WHAT IS BEING COUNTED

At first thought it seems obvious what an accident is, for we have seen them. On further reflection this certitude fades. If a bus rider is injured when the bus stops suddenly, if two cars collide without visible damage, or if someone has a finger amputated by a closing car door, are these 'accidents'? The words 'accident' or 'crash' are given meaning by special purpose definitions and by forms on which police officers and others record accidents. Someone has to decide which events are to be recorded as accidents, and thereby qualify for entry into official statistics. How many of the events that should be recorded are actually recorded is a separate question.

a. What is an accident?

We are interested here in the safety of the road transport system. Thus, collisions, crashes or accidents involving motor vehicles are to be counted. But Agran and Dunkle (1985) speak also about 'non-crash events'. These are the result of a vehicle manoeuvre (sudden stop, turn or swerve) or of some passenger act (falling out of a vehicle, loss of balance) in which injury occurs but in which there is no vehicle crash or impact. Another possible blind-spot are the falls to pedestrians or accidents to bicyclists or in-line skaters which do not involve a motor vehicle. All these seem to be part of the safety cost that goes with mobility on the road but they seldom count as accidents.

Much also depends on who does the counting; the police, the insurance company or the hospital. Each agency has its own purpose, process and definitions to suit. The police record 'reportable motor vehicle accidents', a clearly defined category of events. In engineering studies, accident counts produced by the police are the most commonly used ones. The same is true also for most studies that are not 'engineering' in kind. This is why the focus here will be on reportable motor vehicle accidents. In Ontario,

a Motor Vehicle Accident is "Any incident in which bodily injury or damage to property is sustained as a result of the movement of a motor vehicle, or of its load while the motor vehicle is in motion."

Perhaps more to the point is a definition abstracted from the Ontario accident report form. Here,

a Motor Vehicle Accident is a collision of a motor vehicle with a moveable or fixed object, or its explosion, submersion or rollover.

Both statements require a definition of what a 'motor vehicle' is. A few years back, in Ontario, this category did not include bicycles, vehicles on rails, snowmobiles or tractors. So, if a bicyclist or a pedestrian was killed by a street car, such an event may not have entered the count of motor vehicle accidents. But now, paradoxically, a bicycle propelled by pedal power is legally a motor vehicle.

b. The question of reportability

In Ontario, as elsewhere, not all motor vehicle accidents are reportable; only those which either exceed a specified minimum damage to property or involve an injury. Of course, neither the police officer nor the involved parties can well estimate the amount of property damage. In Ontario, the term 'injury' means either that it is visible or that a person complains of being injured. A fatal injury is when a person dies as a consequence of an injury sustained in a motor vehicle accident within 30 days of its occurrence[1]. Figure 4.1 serves to show that even though reportable motor vehicle accidents are clearly defined, they are still a rubbery yardstick to measure safety with.

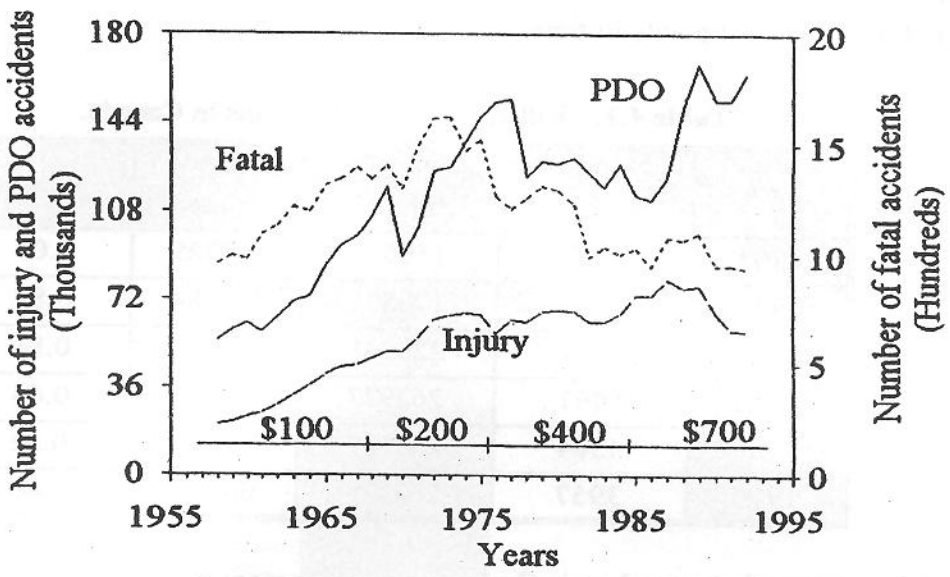


Figure 4.1. Fatal, non-fatal injury and property damage only accidents in Ontario.

Note that because the cost of automobile repairs in Ontario has been increasing, the amount of property damage that makes an accident reportable had to be periodically adjusted as is shown above the horizontal axis in Figure 4.1[2]. The precipitous drops in the counts of property-damage-only (PDO) accidents which correspond to the timing of these adjustments are evident. It follows, that the upward sloping segments in the count of property-damage-only accidents cannot be trusted either. If the cost of repairs increases while the amount that makes an accident 'reportable' remains fixed, an ever larger proportion of accidents becomes reportable. Therefore, an increasing count of PDO accidents may reflect, in part, an erosion in the value of money. But cost of car repairs or the rate of inflation have little to do with safety. Thus, if one were to use the count of property-damage-only accidents to assess the effect of some treatment, without taking into account the accident inflation as caused by monetary inflation, one might come to incorrect conclusions.

Many countries keep only records of injury accidents. The count of injury accidents does not depend on the cost of car repairs nor on the value of money. Would then the count of injury

---

[1] In Quebec it is eight days, in Prince Edward Island 12 months.

[2] In 1990, when in Ontario the reportability limit was $700, in Prince Edward Island it was $1000 while in New Brunswick and in British Columbia it was $400.

accidents obviate the difficulties which beset the count of property-damage-only accidents? The count of injury accidents and of fatal accidents is also shown in Figure 4.1. Note that the two follow very different time trends. Consider the data in Table 4.1 that lists the number of persons killed and injured in Canada for six points in time.

Table 4.1. Killed and injured persons in Canada.

Year	Number Killed	Number Injured	Killed/ Injured	Index
1965	4902	150612	0.0325	1.00
1970	5080	178501	0.0284	0.87
1975	6061	220941	0.0274	0.84
1980	5461	262977	0.0208	0.64
1985	4364	259189	0.0168	0.52
1990	3957	263139	0.0150	0.46

It is apparent that nowadays in Canada only half as many persons die per injury as did twenty years ago. A similar pattern is shown for California in Figure 4.2 and has been noted in almost all provinces, states and most countries. The plus signs show the number of persons killed per person injured. The squares show the number of fatal accidents per non-fatal injury accident.

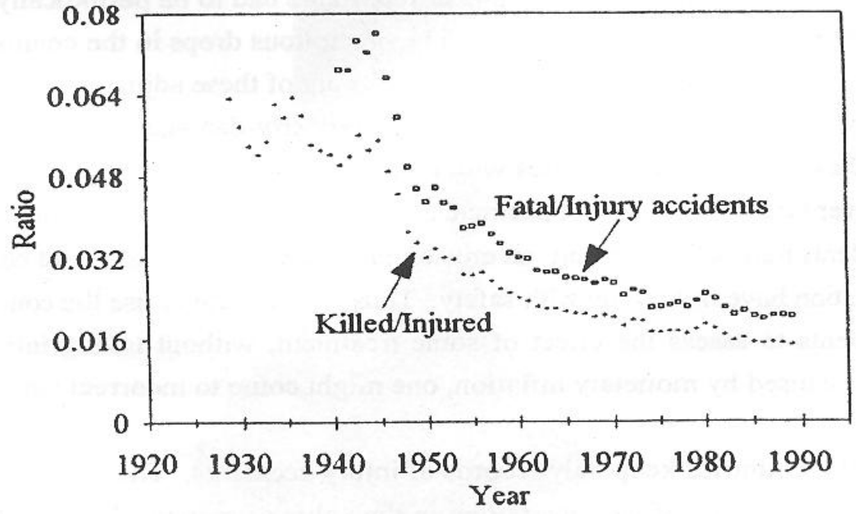


Figure 4.2. The declining odds of dying from an injury accident in California.

This trend may be due to less injurious vehicles and roadside, seat-belt and helmet wearing, more accidents in urban areas, etc. However, the decline may be partly due to an increased inclination to complain of injury. Indeed, in Ontario proportionately more slight injuries are now reported than earlier. Alternatively, perhaps people are more resilient today than 20 years ago, or perhaps the emergency medical services are better. In addition, we now have better medicine and more people are kept alive beyond the 30 days after the accident (thereby removing some victims from the count of motor vehicle accident fatalities in Ontario). All these factors need to be recognized. In short, a change in the count of injury and fatal injury accidents may be the result of a variety of factors. Among those are factors such as the changing cost of car repairs, better medicine, urbanization, aging of population, advances in nutrition, people's inclination to sue and the like.

Do we intend to say that if people complain more readily of being injured, the road transport system is less safe? Probably not. Do we intend to claim that if medicine improves, or if more money is available to keep people alive, roads are safer? Perhaps we do, but this is a qualified yes. In any case, the contention is, that even the count of fatalities and injuries is a somewhat rubbery yardstick. Although the temptation to attribute observed changes in accident frequency to the treatments we are responsible for is strong, one must not assume that without these treatments accident frequencies would have remained constant. The existence of time-trends needs to be recognized in our analyses. (How this can be done will be discussed in Chapter 12.)

We must also keep in mind that our raw material is accidents made reportable by specific criteria for reportability. To illustrate, assume that the chances of 70 years old pedestrians and 20 years old pedestrians to be struck by a vehicle are the same. When struck, many older pedestrians are injured and die, while most young pedestrians survive, some even walk away without leaving a trace in the official accident records. When later the police statistics are examined, older pedestrians will be found 'over-represented'. It will look as if they have more injury and fatal accidents than young people. The 'obvious' explanation of this phenomenon will be sought in deteriorating vision, in longer reaction times and in slower cognitive functions. Bureaucracies will sprout and 'programs' aimed at older road users will abound. But the premise of this thought experiment was that the chances of young and old to be struck by a vehicle are the same. This statistical mirage of over-representation contributes to the common opinion (prejudice?) that the elderly have more accidents due to some deterioration of the senses and agility which go with aging. As a result of the misidentification of the cause of the over-representation, it became fashionable to ask the elderly to make vision tests a condition for license renewal, to reexamine their knowledge of traffic laws, and to think of other ways to limit their driving privileges. However, the way this example was built, they do not have a problem of cognition, only of brittle bones. The cure for brittle bones is not periodic vision testing nor license revocation. The elderly are hardly a menace to others (see, e.g., Evans 1991) and the regulatory attention they are now getting is difficult to justify.

c. Incomplete reporting

Not all accidents that are reportable are also reported. Hauer and Hakkert (1989) reviewed 18 studies in which researchers examined police, hospital and insurance sources for common entries. They conclude that "the police miss some 20% of injuries that require hospitalization and perhaps up to half of the injuries that do not." Perhaps 60% of reportable property-damage-only accidents are not reported. The inclination to report an accident to the police was found to increase with the age of the injured person, and with the number of vehicles involved. Injuries to non-occupants were less completely reported than injuries to passengers and these, in turn, less than injuries to drivers. Thus, not only is the count of accidents only a partial one, it is not uniform for all accidents. This makes the interpretation of accident counts difficult.

To illustrate, consider a road system for which the 1994 official records show 50 single vehicle accidents in which the driver was young and impaired (Table 4.2). If only one in four such accidents is reported, as is likely, about 200 such accidents have occurred. Few would disagree that it is the 200 accidents that should guide safety management, not the 50 that got reported. Still most statements about road safety which one reads and hears are about the (50) accidents reported, not about the (200) accidents occurring.

Table 4.2. Illustration.

	Number of reportable accidents	Proportion of accidents reported	Number of reported accidents
1994	200	1 in 4	50
1995	200	1 in 5	40

Imagine now that at the end of 1994 stiffer penalties for drinking and driving were introduced. Assume that the number of reportable accidents remained unchanged, but the inclination to report them changed from one in four to one in five. Since we only know about accidents that are reported, ostensibly, between 1994 and 1995 there was a reduction of 20% in such accidents. If this were a correct estimate then saying that the reduction was of 10 accidents would be wrong; one should claim a reduction of 40 accidents (20% of to 200). Thus, the partial reporting of accidents leads to an underestimate of the size of the problem, and also to an underestimate of the effect of treatments. However, the insidious aspect of partial reporting is that if the inclination to report accidents changed, one cannot draw reliable conclusions from changes in the number of reported accidents. In this case, our impression of a 20% reduction is false; it is entirely due to the changed inclination to report the occurrence of the event to the police.

Too little is known about the extent of under-reporting and changes therein. The situation reminds one of Plato's cave, in which only the shadows of real events are visible. What we observe are **reported accidents** while what we wish to count are the **accidents that occur**. We see a subset of what we wish to see but how much is hidden from view is not known.

It is commonly assumed that the difference between the number of accidents reported and the number of accidents occurring is unimportant if one is interested in the ratio of before and after accidents because the proportionality factor cancels out. This assumption is true only if the proportion of reported accidents is precisely the same in the 'before' and the 'after' periods. In the previous example in which the proportion of reported accidents declined from 'before' to 'after', the ratio 40/50=0.8 still gives the incorrect message of a 20% reduction $[(1-0.8)\times 100]$. For statistical details see Hauer and Hakkert (1989).

The uncertainties of counting, of classification, and of partial reporting inhibit understanding and action. To see how, consider a study, perhaps the best of its kind at the time it was done (Zegeer et al., 1987), in which accident data from seven states in the U.S. have been pooled to model the safety of rural roads as a function of their traffic, geometry and roadside. It turned out, that under similar conditions of traffic and geometry, perhaps four times as many single-vehicle accidents are reported to occur per mile of rural two-lane road in Alabama as in West Virginia. This is partly due to differences in accident reporting. Imagine now that roads in Alabama have, say, somewhat worse geometric features than roads in West Virginia. The statistical software used information about, say, lane width and shoulder type, but does not know about the differences in accident reporting. Therefore, it will attribute the higher accident frequency in Alabama to lesser lane width, unpaved shoulders, etc. What in reality is a difference in accident reporting, might appear to the unwary analyst as a relationship with road geometry. When the results are later published, conscientious engineers change the design standards and these shape the design of future roads. Obviously, when rates of reporting differ, data about accident counts cannot be simply pooled[1].

If data cannot be simply pooled, is it the upshot that each state, perhaps each county, city, or neighborhood has to be viewed as a separate case? May one reasonably claim, as many are wont to do, that: "Our conditions are different, and therefore their results do not apply here?" The tendency to so claim is a paralyzing affliction. It evokes the memory of the middle ages, in which the boundaries of the universe of knowledge were often the boundaries of each fief. When data have

---

[1] To come to more defensible conclusions, the authors could have examined whether the same relationships with lane width, shoulder width etc. hold in each of the seven states.

only local validity, discovery and generalization cannot prosper. In recognizing the real differences between jurisdictions one must also see the overriding similarities in motor vehicles, roads and human beings. These similarities provide a solid basis for the hope that findings from one jurisdiction are likely to be relevant to action in another. This issue will be revisited in a quantitative way in Section 10.

In summary, it must be remembered that the number of accidents reported is less than the number of accidents occurring and that the ratio of the two is largely unknown. For the conduct of a Before-After study what matters is whether there was a change in the ratio (accidents reported)/(accidents occurring) and, if yes, how to account for it.

d. Errors

Once an accident is reported and the report form is filled out, it goes through a process of pre-coding, coding, entry, computer editing and retrieving. Errors creep in at all stages of this process. Some are impossible to detect. Others can be examined by comparing the original report form to the computer printout.

In one such examination (Kelman, 1977), 4.6% of the codes of an average record had some error in them. The most common errors had to do with the residence of the driver, type of vehicle, impact area, location on roadway, road character, direction of travel, apparent driver action, pedestrian condition and charges. In another study, Shinar et al. (1981) conclude that the accident variables least reliably reported by the police were : " . . . vertical road character, accident severity and road surface[1]. The most reliably reported data were those of general location, date, number of drivers, passengers and vehicles. The informativeness of the police reports with respect to driver /vehicle characteristics was practically nil, with the exception of driver age, sex and vehicle model."

For treatments that focus on specific sites, most serious inaccuracies are those of location. It is easy for a police officer to record Street as Road or an accident on a freeway ramp as having occurred 'mid-block'. When the analyst later searches the data base from accidents occurring on Main Street or on freeway ramps, such records are almost irretrievably lost. Janusz (1995) found that some 40% of accidents on freeway deceleration lanes in Toronto were coded as mid-block. In another study, the same street name was coded in seven different ways. Thus, the last contribution to the uncertain nature of the data are errors in filling out the accident report form and errors in its processing.

---

[1] Accident severity was wrong in 30% of the cases, speed limit in 40%, vertical character in 41%, defective eyesight in 42% etc.

e. Summary

The most common data in the study of road safety are police-reported accidents. It follows that statements about safety which are made on the basis of police-reported accidents are influenced by the definitions, forms, police procedures, reportability criteria, inclinations to report, and errors. The recognition that the counts of reportable accidents are not the hard evidence one usually associates with numbers is humbling. The pessimist would be justified in asking: "should we proceed? Is statistical analysis meaningful?" There is no convincing way to counter this observation; one can only soften its impact.

To begin with, expectations should be realistic. Our work is not done in the laboratory and we cannot control for all important causal factors while changing one factor at will. Therefore, the standards of confirmation or refutation that apply in physics or chemistry are not within reach. Nor can we aspire to approach the standards of research in medicine, experimental psychology or agriculture, as long as we are denied the ability to conduct randomized statistical experiments. Thus, we can neither arrest change, as in the laboratory, nor neutralize its effect by randomization. Therefore, the uncertainties of accident count data must not be viewed as compromising the high standards attainable in the exact sciences. In an observational study of road safety, the uncertainty surrounding accident counts is but one of the problems that need to be recognized and, if possible, accounted for.

In addition, it seems fair to note that with all their shortcomings, databases containing reported motor vehicle accidents are uniquely rich and complete in comparison to data bases available to those who study industrial and occupational safety, and often even medicine. It is a stroke of good fortune that data for road safety research are permanently and routinely collected in a standardized form by well-trained personnel - the police. There are also special purpose data bases such as FARS (the Fatal Accident Recording System) and NASS (the National Accident Sampling System) as well as data from hospital records and occasionally, insurance companies.

For all these reasons it would be wrong to conclude that progress to better road safety knowledge is impossible. What one may conclude is that to make progress is far from easy; that the methods used should match the nature of the material to which they apply; that one must admit and clearly state the limitations stemming from the material and circumstance.

4.2 TARGET ACCIDENTS

In Chapter 2 I mentioned the R.I.D.E.[1] enforcement program. The aim of R.I.D.E. was to reduce driving under the influence of alcohol. Accordingly, the safety effect of R.I.D.E. was sought in 'alcohol-related injuries'. These were the **target accidents**. The identification of target accidents is not always a simple or clear-cut matter. It will turn out that the effect of R.I.D.E. should not have been sought in alcohol-related-injuries. Similarly, I have already mentioned that while by road illumination one expects to affect mainly nighttime accidents, changes may occur in daytime accidents as well, due to the presence of new lampposts and the guardrail in front of these. Thus, thinking about target accidents requires that one be able to answer the question:

"The occurrence of what kind of accident will be affected by this treatment?"

A good answer demands an understanding of the **accident generation process** - how will the treatment work. In endnote 'a' of Chapter 2 I list the rules for judging causality in epidemiology. Rule number two demands that there be a link between 'dose' and 'response'; rule number five asks for biological plausibility. Taken together, these rules require that there be some understanding how and why a causal factor can have an effect. The same principles apply in road safety. Thus, one can state that:

The **target accidents** of a treatment are those accident types the occurrence of which can be materially affected by the treatment.

To illustrate the difficulties of deciding which accidents types are the target of a treatment, here are a few stories.

Story 1.

The reader may have accepted, as I did for a long time, that alcohol-related injuries are legitimate 'targets' for a treatment aimed at reducing alcohol-impaired driving, such as R.I.D.E.. However, a moment's reflection will cast doubt on this legitimacy. For the sake of the argument, assume that without R.I.D.E. a person drives 10,000 km/year sober, and 5,000 km/year with some alcohol in the blood; with R.I.D.E. this would change to 12,000 and 3,000 km/year, respectively. The total distance traveled remained the same, only the proportions with and without alcohol in the bloodstream changed. In consequence, R.I.D.E. (if successful) causes a

[1] Reduce Impaired Driving Everywhere.

reduction in alcohol-related injuries but also an increase in non-alcohol-related injuries. To consider only alcohol-related injuries as 'target' leads to a bias favoring this kind of treatment.

By the criterion that target accidents are those the occurrence of which can be materially affected by the treatment, non-alcohol-related injuries are also the target of this kind of treatment. The change in non-alcohol-related injuries should too have been considered in estimating the effect of R.I.D.E. on safety.

Story 2.

A state highway agency was building sound-walls alongside some of its roads. (A sound-wall is a structure located well off the traveled way and intended to shield people living and working near the road from traffic noise). As is true for any solid object near the road, sound-walls can be collided with. Because some of the money came from the federal government in Washington, the agency had to determine the impact of the sound-walls on safety and report back. The analyst chose to examine the change **in total accidents** from the 'before-sound-wall' to the 'after-sound-wall' period. Washington did not like the analysis, arguing that sound-walls could hardly be expected to influence collisions between vehicles which never left the road.

If Washington was right, one should look for the effect of sound-walls on safety only among the ran-off-the-road accidents. Perhaps one would have to further restrict this to those 'before' accidents where the vehicle ran off the road and strayed beyond where the sound-wall now stands. This would have to be compared to collisions with the sound-wall during the after period.

The state highway agency and its analyst disagreed with Washington, arguing that sound-walls may act as visual delineators, thereby reducing ran-off-the-road accidents in general. Furthermore, it is possible that the presence of sound-walls reduces speed and this, in turn, may have an effect on all accidents.

The debate cannot be resolved by exchanging memoranda. The question is empirical: does the presence of a sound-wall at distance X from the edge of the pavement affect driver behavior in terms of speed choice, position within lane etc.? This question can perhaps be answered, but it may cost more to do so than was the cost of building the sound-wall. The state highway agency and its analysts may be right. However, unless they can give proof that there exists a process and mechanism by which the presence of the sound-wall materially affects the probability to run off the road or alters the behavior and interaction on the road, Washington's position looks the more tenable to me.

The main merit of this story is not in emphasizing the difficulties of deciding what is 'target' and what not. Its essence is in the trade off between statistical precision and the clarity with which the impact of the treatment can be traced. If only collisions with the sound-wall are defined to be 'target', the accident count is likely to be small, perhaps too small to say anything with statistical precision, but the link to the treatment is clear - only the presence of the sound-wall could be responsible for such collisions. If total accidents are considered to be 'target' the accident counts are more numerous. However, the link with the treatment is fuzzy - surely many causal factors may be responsible for the change. Without accounting for their effect, it is impossible to isolate the effect of the sound-wall.

Story 3.

> Story 2 was about a cause and the uncertainty about what accidents it will affect. Story 3 is about an opposite case: there was an effect but it was unclear what its cause might have been. I have been asked to examine the safety effect of two kinds of resurfacing projects on two-lane rural roads in New York State. I found that in projects consisting of only putting a new blacktop on the road there was a large adverse effect on non-intersection accidents and a lesser adverse effect on intersection accidents. In projects where, on the occasion of resurfacing, various safety-related improvements were made, non-intersection accidents remained stable while intersection accidents were drastically reduced. Are both non-intersection accidents and intersection accidents the 'target' of resurfacing?

Earlier research hinted that a new blacktop increases speed and that higher speeds will tend to result in more severe consequences of collisions. Thus, it is not difficult to think of a mechanism by which a new blacktop might cause an increase in reportable non-intersection accidents. One might also be willing to believe that the various safety related improvements could offset the adverse consequences of higher speeds for non-intersection accidents. But how exactly did the reduction in intersection accidents come about, when the safety-related improvements at intersections were only the replacement of missing signs and the clearing of brush? We have found an effect but do not have a clear idea how it could have come about.

In principle, the change in the frequency or severity of target accidents is the **product** of a treatment. Target accidents are those that can be substantially affected by a treatment. The recognition of the group of accidents that may be affected by a specific treatment requires knowledge about the **process** of accident generation. Knowledge of the process is essential to estimate the product of a treatment. It takes expertise to understand the process of accident generation. However, even experts can be surprised as shown by the following story.

Story 4.

The use of daytime running lights was first mandated in Finland. It was preceded by elaborate laboratory research on how this might improve the detection of oncoming vehicles by drivers. The Finnish law required that lights be used during the winter outside of built-up areas. The assumption must have been that this is where using headlights during the day will do most good. Sweden followed later but its law required the use of daytime running lights in summer as in winter and in cities as in the countryside. When the effect of the Swedish treatment was later evaluated (Andersson and Nilsson, 1981) it turned out that of the estimated reduction of 2230 injury accidents, 1524 were during the summer and in built-up areas. Unexpectedly, pedestrians and bicyclists were important beneficiaries.

Thus, expert understanding of the process of accident generation is at times limited. The lessons of hindsight need to be followed up.

An additional aspect of this experience is of interest. Andersson and Nilsson argued that nighttime accidents could not be affected by the use of headlights during the day and used these as 'comparison' accidents. Comparison accidents are used in Before-After studies to predict what would have been the number of target accidents had the treatment not been implemented. I will devote a full chapter to this subject. Here I only note that the notions of 'target' and 'comparison' accidents are closely linked. If 'target' are those accidents that can be materially affected by a treatment,

> **comparison accidents** for a treatment are those accidents the occurrence of which cannot be materially affected by the treatment.

Naturally, just as it is difficult to judge what accidents are 'target' so it is difficult to be convincing about which are 'comparison'. This source of an annoying ambiguity is further illustrated by the following episode.

Story 5.

To estimate the effect of allowing vehicles to turn right during the red phase of a traffic signal, accidents at signalized intersections in Alabama and South Carolina were divided into two groups (Clark et. al., 1984). Target accidents were those involving at least one right-turning vehicle; comparison accidents were those in which none of the colliding vehicles turned right. Counts of accidents before and after right-turn-on-red was allowed were used. To illustrate the logic of the analysis, consider one of their tables in which the following accident counts were given.

Table 4.3. Right-turn-on-red (Clark et al., 1984.)

	Target	Comparison
Before	167	3566
After	313	6121

The usual assumption is that, were the treatment (here, allowing right-turn-on-red) not implemented, the change in the target and comparison accidents would be the same. If so, the 167 'before' accidents would turn into 167x6121/3566= 287 'after' accidents. The difference 313-287 can than be attributed to the treatment.

The safety consequences of allowing right-turn-on-red were subject to heated controversy. To support its policy of allowing right-turn-on-red the Institute of Transportation Engineers supported the conclusions of a study which was similar to that shown in Table 4.3, but with a twist. The reconstructed accident counts are given in Table 4.4.

Table 4.4. Right-turn-on-red (Hooper, 1981, reconstructed.)

	Right-turn	Other
Before	2192	28656
After	2808	26344

Consider, as before, the 'right-turn' accidents to be 'target' and the 'other' accidents to be 'comparison'. If so, 2191x26344/28656=2014 right-turn accidents would be expected during the after period were right-turn-on-red not allowed. The indication would be, that there was a 28% increase in target accidents attributable to allowing right-turn-on-red, a conclusion that did not fit the current ITE policy. However, if all intersection accidents were to be considered 'target', then in the 'before' period there were 30848 accidents and in the after period 29152, a reduction of 5.5%. This fit the policy better, and was endorsed (Hooper 1981).

It appears then, that whether one conclusion is drawn from a data-set or its opposite, depends on which accidents are considered the target of the treatment. Thus, the task of identifying target and comparison accidents properly overshadows in its importance the statistical questions to which much of the later chapters are devoted. I advocate the view that target accidents are those that can be affected by the treatment. It is clear that accidents involving vehicles turning right against red are target accidents. It is also plausible that other types of accidents may be affected. Thus, e.g., if more vehicles turn right on red, fewer may turn right on green. If so, the latter category are also target accidents.

One could perhaps argue that allowing vehicles to turn right on red so thoroughly alters the operation of a signalized intersection that all types of intersection accidents can be affected. I think that this is far from self-evident. Accordingly, if a person chooses to search for the effect of allowing right-turn-on-red in, say, accidents to straight-through vehicles and left-turning vehicles, the onus is on that person to show by what **process** this influence can materialize, and to support this contention (that the process is indeed associated with the treatment) by field data. If one cannot show that allowing right-turn-on-red brings about a process that materially affects accidents to through-vehicles and left-turning, these are not target accidents.

Failure to identify all accidents that are the 'target' of a treatment is in the background of a paradoxical result elucidated by Griffin (1990 a). He writes about studies of the safety effect of seat belts (e.g., by Tourin and Garet, 1960, and by Campbell, 1968). In these studies conclusions are drawn from the differences in the proportions of injuries among belted and unbelted drivers.

Story 6.

For clarity, Griffin uses the following hypothetical example. Suppose that we have information about the injury severity of 1000 belted and 2000 unbelted drivers who were involved in accidents:

Table 4.5. Hypothetical data.

Driver Injury	Unbelted	Belted
Fatal	20	6
A-level[1]	80	24
B-level	120	36
C-level	240	72
None	1540	862
Totals	2000	1000

On this basis it appears that seat-belt wearing reduces the probability of a driver being injured in an accident by 40%. The argument goes as follows. If seat-belt wearing did not affect the chance of being killed, then, of the 1000 belted drivers in accidents one would expect 1000×(20/2000)=10 to be killed. Since not 10 but 6 were killed, this indicates a 40%

[1] Non-fatal injury accidents are judge as 'incapacitating', 'non-incapacitating' and 'possible injury'.

reduction in the probability of being killed in an accident. The reader can confirm, that by the same reasoning, the 40% reduction obtains for the other three injury levels as well. Suppose now that, as is common in many countries, no information exists about accidents where there is no injury. That is, the 'None' row in Table 4.5 does not exist and the totals in the last row are 460 and 138. Using the same procedure as before, among the 138 belted drivers one would expect to find $138\times(20/460)=6$ drivers killed. Since 6 were killed, seat-belt wearing now appears to have no effect. The reader can again check that the same conclusion holds for the other three injury levels. Similarly, one can show that if, say, all injury accidents were reported but only 50% of the no-injury accidents, the effect of seat-belts would appear to be a 35% reduction. What is happening?

The premise of the example is that wearing a seat-belt may affect the probability of sustaining an injury. That is, each accident is considered to be a 'trial'. The outcome of such a trial may be 'fatal', 'A', 'B', 'C' or 'none'. In our computation we used the factor 20/2000 as an estimate of the probability to be fatally injured in an accident. But this is a sensible estimate only if the 2000 is the correct count of the number of trials. If for whatever reason we do not know the number of 'trials', that is the total number of drivers in accidents in the last row is not known, then no statement about probabilities or changes therein can be made. In our case, accidents to unbelted drivers are the 'target' of the treatment. An incomplete count of these is seen to lead to incorrect inferences.

Other stories.

That the effect of a treatment is difficult to anticipate and trace is well illustrated by 'accident migration'. I hasten to say that by 'accident migration' I do not mean that drivers seek some constant level of risk, and that if some facility is made safer they will compensate by riskier behavior elsewhere. What I mean is that even localized treatments may have non-localized consequences. One common consequence of treatment is traffic diversion. Clearly if treatment here causes changes in traffic elsewhere, safety elsewhere will change too. However, accident migration can take more subtle forms. Thus, e.g., I live in a neighborhood where almost all intersections have been converted from two-way to all-way STOP control. Therefore, at the few intersections which remain under two-way STOP control, I catch myself making the mistake of expecting vehicles on all approaches to stop. For me, and perhaps for others, the intersection that remains two-way STOP controlled is now more dangerous. This is a plausible mechanism for accidents migrating from intersections converted to all-way STOP control to those which remain unconverted. Such accident migration has been examined by Persaud (1986, 1987). Similarly, Gårder believes that the effect of 'continuous shoulder rumble strips' should not be sought only in the reduction of single-vehicle-run-off-the -road accidents. A drowsy driver may be woken up by an incursion on the rumble strip. But since drowsy drivers tend to doze off again, the next event may be a multi-vehicle accident. Thus a single-vehicle accident prevented may materialize as a multivehicle-accident created.

It is clear that the question, "which are the target accidents for a treatment?", is of utmost importance. Target accidents are all those that can be affected by the treatment. Thus, there must be a causal link between the treatment and its target accidents. At times the link is self-evident[1]. Also, based on a current understanding of the process by which accidents are generated one can often judge the presence of a causal link. It is more difficult, perhaps impossible, to judge the absence of a causal link. Therefore, one cannot say with conviction that a certain class of accident is definitely not affected by the treatment. In my view, when the causal link is questionable and far fetched, one should not consider such accidents as target (unless evidence has been produced about the process by which the treatment influences these accidents). Regrettably, I cannot give more definitive guidance on the matter. Better guidance would be possible, if more was known about the process by which accidents are generated and avoided. Till such an understanding is reached, some ambiguity will remain. It follows that a convincing evaluation of the safety effect of a treatment requires a good understanding of the **process,** not only an estimation of the **product.**

4.3 CHAPTER SUMMARY

Accident counts are the raw material from which treatment evaluations are fashioned. Two important aspects of accident counts were examined in this chapter. The first aspect is that of reliability and accuracy. What we call 'accident' or 'crash' is a matter of definition. This definition prescribes what should be counted. The definition tends to change with time, differ from place to place, and depends on the agency collecting the information. This restricts our ability to pool data and to make comparisons between times and places. Not all reportable accidents are reported. Only an uncertain fraction of what is occurring finds its way into the official record. The proportion of reportable accidents that are reported depends on age, gender, weather, number of vehicles involved, insurance procedures, police practices and a host of other factors. In addition there are the errors and inaccuracies that plague the process by which a reported event is transformed into coded information. All this should instill a corresponding amount of humility and caution that is needed to counteract the ill-founded confidence that a printed number tends to inspire. Nevertheless, this realistic assessment of the limitations of our raw material is no ground for pessimism about the possibility to create factual knowledge about the safety effect of treatments. Our discipline is more

---

[1] As Story 1 shows, even when a link appears self-evident, it may not be so.

fortunate than others in that data are permanently and routinely collected for us in standardized form by trained police officers.

The second important issue discussed in this chapter is the difficulty of deciding what accidents are the target of a treatment. This is logically entangled with the important question of which accidents are not affected by a treatment. I do not think that the discussion here has brought the matter to a satisfactory closure. The diverse problems were illumined by six stories which, unfortunately, are not invented ones. The logic is simple enough. Target accidents are all those that can be affected by the treatment, comparison accidents are those that cannot be so affected. The problem is that to decide whether some accident class can be affected by a treatment is not easy. It demands an understanding of the accident generation process - the complex causal web in which treatments affect nature, both human and inanimate.

At present, our understanding of the accident process is seldom equal to the task. Perhaps it will improve. On the other hand, perhaps it is unrealistic to expect that, in a world in which all may affect everything else, it is at all possible to satisfactorily label some accidents 'target' and others as 'comparison'. Be it as it may, the six stories serve to point out that important questions surrounding the interpretation of observational studies are not statistical in nature. What result one obtains from a study depends less on uncertainty that is due to randomness and more on non-statistical decisions about what is 'target' and what is 'comparison'. What practical guidance can one offer in this situation? It might be useful to think of the three accident groups in Figure 4.3.

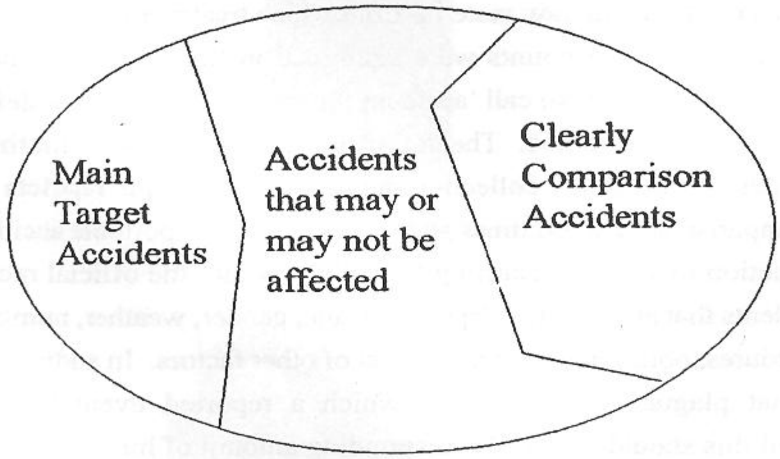


Figure 4.3. Three groups of accidents.

It is usually possible to recognize accidents that are the 'main target' of a treatment. To do so may not always be as easy as it seems, but is essential. However, there are practical limits to such an effort. An attempt to broaden the aim of a study to encompass all possible accidents that could be affected will often make the results too diffuse to be recognized and too uncertain to be of interest. Thus, it is sensible to confine observational studies to 'main target accidents'. This carries with it the obligation to state clearly which other accident groups could have been affected but were not examined.

A comparison group of accidents, to be of any use at all, must be clean; the possibility of the treatment to affect their occurrence or non-occurrence must be extremely remote. In some studies it is possible to count the comparison group accidents on entities that were not treated. If so, it may be sufficient to ensure that traffic diversion or altered road user expectations as caused by the treatment could not have materially affected the safety of entities in the comparison group. In other studies the comparison group of accidents is for the treated entities[1]. Since causal links in this world are many, complex and often unexpected, the peril of seeing mirages as treatment effect is especially acute in this case.

---

[1] For example, if the treatment is equipping vehicles with Daytime Running Ligths and nighttime accidents serve as 'comparison'. Similarly, if the treatment is the enactment of a seat belt law and injuries to unbelted passengers in the same vehilce serve as 'comparison' when the effect of seat belt use for drivers is studied. Also, if the treatment is allowing Right-Turn-On-Red at intersections and accidents to vehicles not turning right serve for 'comparison'.

CHAPTER 5

PREDICTION AND ESTIMATION

At this point it would be possible to plunge into the analysis of the various Before-After study variants. However, the many options, variants and corrections to be discussed in the next few chapters might be bewildering without a sense of purpose and direction. The aim of this chapter is to show that the many ways of doing a Before-After study merely reflect the many possible ways for accomplishing two basic tasks:

1. The task of **predicting** what would have been the safety of the entity in the 'after' period had treatment not been applied, and

2. The task of **estimating** what the safety of the treated entity in the 'after' period was.

As stated in Chapter 2, and shown in Figure 2.2, the effect of a treatment on the safety of some entity consists of comparing the prediction of what safety would have been to the estimate of what safety was. The preparation of numerical values for the prediction and the estimate are the two basic tasks that are common to all studies of the Before-After kind.

5.1 PREDICTION OF WHAT SAFETY WOULD HAVE BEEN

To show that there are many ways to predict, consider again the R.I.D.E. enforcement program described in Chapter 2 and the injury counts[1] shown in Figure 5.1. The + signs are the counts of injuries sustained in alcohol-related accidents in each of five years before the implementation of R.I.D.E.. From these, and perhaps from the knowledge of other relevant facts, we need to construct a prediction of what would have been the expected number of injuries during the 'after' period had R.I.D.E. not been implemented. Because the safety effect of a treatment may

[1] As argued in Chapter 4, both alcohol-related and also non-alcohol related injuries are 'target'. In the illustrations to follow, attention is confined to the prediction of alcohol-related injuries only.

vary with time, a prediction is needed for several periods after treatment. In Figure 5.1, for simplicity, a prediction for only one year is shown.

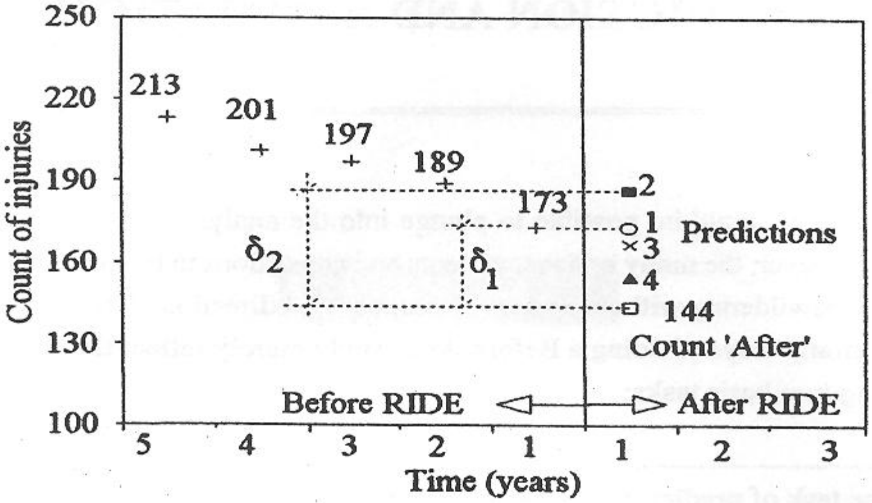


Figure 5.1. Injury counts, 'predictions' and 'effects'.

There are many ways to predict, and, accordingly, many different predictions. One such prediction is shown in Figure 5.1 by the circle. Note that it is the same as the count of injuries one year before R.I.D.E.. This prediction is based on the belief that what happened one year before R.I.D.E. would have happened one year later, had R.I.D.E. not been implemented. Looking at the trend of the + signs, this seems to be a peculiar belief and strange way to predict. Yet this is what one would normally do in a naive Before-After study with a one-year 'before' period.

Another prediction is shown by a black square. It is the average of the last three years of injury counts. The use of such a prediction amounts to a belief that one gains more by 'averaging' than by recognizing the possibility that there is a time-trend beneath the data. It is often taken for granted that a three-year 'before' period will give better a prediction than a one-year 'before' period. However, this is not likely to be true if there is a sizeable underlying time-trend.

The third prediction (shown by a × in Figure 5.1) comes from a linear least-square fit to the five years of injury counts. This is a sensible way of predicting here although one can have little confidence in the extrapolation of a trend based on five points.

The fourth prediction is shown in Figure 5.1 by a triangle. This prediction is computed using a neighboring police district as a 'comparison group', one where R.I.D.E. has not been implemented. In this comparison district there were 289 alcohol-related injuries in the last year 'before' period and

259 during one year 'after'. Was the 'treatment' district to undergo changes similar to the 'comparison' district, one could expect $173\times259/289=155$ injuries to occur in the 'treatment' district if R.I.D.E. was not implemented. This is the prediction shown as a triangle. It represents what is typically done in observational Before-After studies in which a comparison group is used.

The four simple-minded predictions shown in Figure 5.1 were obtained by methods commonly used in the study of 'Before-After' data. Predictions 1 and 2 are probably a continuation of the science-laboratory frame of mind mentioned in Chapter 2. The underlying assumption is that during the study period there has been little change in the factors that affect safety. Therefore the safety of the entity under scrutiny is thought to have been constant during the 'before' period and would remain so during the 'after' period, if the entity was not treated. Prediction 3 reflects a more recent tradition. Trust is placed in the extrapolation of a time series of data. It reflects a belief about both nature and knowledge. The belief is that the time trend is the result of many interacting processes that are too complex to measure and understand. Therefore, abandoning all hope to untangle the causal web, one attributes all change to the passage of time. One must also believe that 'natura non facit saltus'[1] and that things will continue into the future much as they have been in the past. Prediction 4 is most likely rooted in the tradition of randomized experiments originating in agriculture and widely used now in medicine and the social sciences. Its hallmark and central idea is the random allocation of entities to the 'treatment' and 'control' groups, a feature almost never found in road safety studies.

One cannot really pretend to believe that all factors influencing safety remain constant from the 'before' to the 'after' period. Therefore, predictions 1 and 2 have no defensible basis. Nor does it seem appropriate to think that the effect of important causal factors (traffic flow, precipitation, driver demography etc.) cannot be measured, understood and modeled. Therefore it is unlikely that the extrapolation of a time trend, as in prediction 3, will produce better results than an attempt to account for change in causal factors by modeling. The use of a comparison group, as in prediction 4, is deficient on two counts. First, just as time-based extrapolation, one uses a comparison group when it is not known what factors affect safety, how they changed and what their effect on safety was. This circumstance is well suited for randomized trails but does not fit well the reality of road safety. In road safety the change in important factors (such as traffic flow or the number of rainy days) is known and can be accounted for explicitly. Second, in road safety studies, comparison groups are usually selected after the treatment group was appointed (and often after the treatment has taken place). Thus, the central tenet of random assignment to treatment is entirely missing.

---

[1] Nature dislikes jumps.

Each of the many possible predictions gives rise to an estimate of change in the number of injuries for the first year after implementation. Two such estimates are shown in Figure 5.1. If the first prediction is used, δ_1 is the estimated change; if the second prediction is used, δ_2 is the estimated change. Each prediction leads to an estimate of change. Because the predictions of 'what would have happened' differ, estimates of the change in injuries are all different.

The existence of many different predictions and the correspondingly many estimates of change in safety is disconcerting. It is one of the reasons why the professional literature abounds with conflicting findings. Which prediction is more valid, which is more reliable? It would be useful to be clear about the merits and shortcomings of each method of prediction so as to be able to judge which results are more trustworthy and what method of prediction is to be preferred in use. While some guidance can be provided, many issues will remain unresolved.

However, a general principle can be stated even now. I think that one should **prefer that method which predicts best**. This is a potentially useful principle. One only needs to apply several contending methods of prediction to the same data, to find out which method would have been best, had it been applied. The results are sometimes revealing and surprising (see Hauer et al.1991, Hauer, 1991b and Quaye and Hauer, 1993). One thing is certain; it is possible to develop methods of prediction that are better than many of those now in use. This task must be tackled quantitatively. I will discuss some improvements to conventional methods of prediction in Part II. In Part III I will suggest some further improvement to methods of prediction.

To predict well one has to answer several strategic questions:

Question a. How to account for those causal factors that **affect safety**, are **measured**, and the **influence of which is or can be known**. Thus, e.g., we know that traffic flow affects safety. For some cases we also have an idea about the relationship between traffic flow and safety. If information about traffic flow both before and after some treatment is available, we need to account for it when attempting to predict what safety would have been in the after period had the treatment not taken place. The approach here is one of modeling.

Question b. How to account for the remaining factors that **affect safety** but which are either **not measured** or the influence of which on safety is **not known**. The influence of these factors can be taken into account by time-based extrapolation and by using comparison groups. The hope is that, as knowledge grows, ever more factors that now need to be accounted for under (b) will be accounted for under (a).

Question c. How to account for **selection bias**. The accident history during the before period is an important clue to what would have been the safety during the after period.

However, the same accident history may also be one of the reasons for which treatment is applied. This makes the prediction subject to a bias which goes under the name: regression to the mean.

Question d. How to account for changes in the **extent of accident reporting**. The degree to which reportable accidents are reported changes from time to time and from place to place. Unless this change is estimated, one cannot separate the effect of the treatment from the effect of changes in accident reporting.

Most of the challenges outlined in these four questions will be faced in the next few chapters.

5.2 ESTIMATION OF WHAT SAFETY WAS AFTER THE TREATMENT

In Section 5.1 we have contemplated the task of predicting what safety in the 'after' period would have been had the treatment not been implemented. Such a prediction needs to be compared to what safety in the 'after' period was with the treatment in place. For the after period we have accident (or injury) counts. These, perhaps with additional information, can be used to estimate what safety during the after period was. Thus, instead of prediction, we face the less problematic task of statistical estimation.

To keep matters simple, in Figure 5.1 I have equated the estimate of safety in the after period with the 'after' period counts. Estimates of change in alcohol-related injuries (δ_1 and δ_2) were shown on this basis. Although this corresponds to what is usually done, it may be better to have a more general picture in mind.

In Figure 5.2, the amounts of change in safety are the differences between the safety predicted for the 'after' years had the treatment not been implemented (the squares), and the estimate of what safety in those years was with the treatment in place (the dots). The squares (the predictions) are here an extrapolation - a linear fit to the 'before' period injury counts (the plus signs of the 'before' period), while the estimates (the dots) are a linear fit to the 'after' period injury counts (the plus signs in the 'after' period). In sum, to estimate the safety effect of a treatment, we require **predictions** of what safety would have been in the after period, had the treatment not been applied and **estimates** of what safety was with the treatment in place. The predictions are based on counts (accidents, injuries etc.) in the 'before' period and adjustments for changes in various causal factors from the 'before' to the 'after' period. The estimates are usually based on the 'after' period counts.

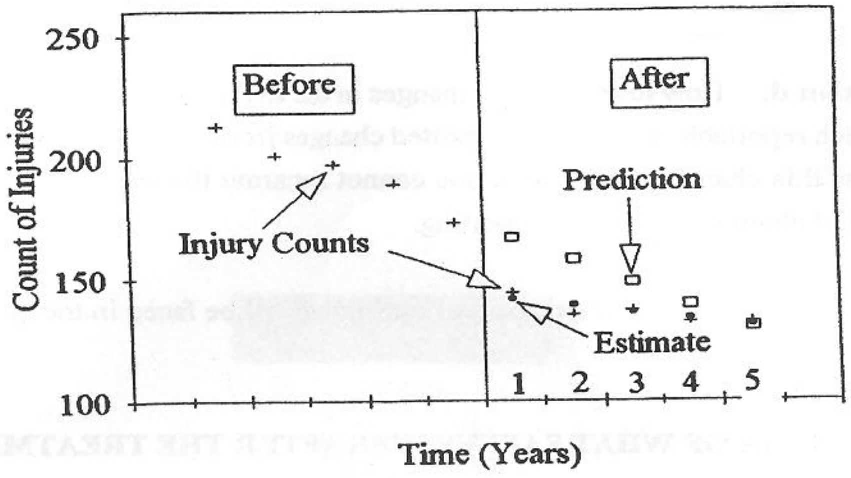


Figure 5.2. Prediction of what would have been and estimation of what was.

5.3 CHAPTER SUMMARY

The estimation of the effect of a treatment on safety always entails a prediction of what safety would have been in the after period had the treatment not been implemented, and the juxtaposition of this prediction to an estimate of what safety in the after period actually was. This simple statement is true for observational studies and also for experiments. The only difference is that in observational studies prediction is more problematic.

The recognition that the central issue is 'how to predict what would have been . . .' has a clarifying, unifying and liberating influence. It clarifies, because one is led to realize what assumptions are made when a certain method of prediction is used. Thus, e.g., if one compares 'before' accidents to 'after' accidents, one predicts by making the (questionable) assumption that, on the average, the count of accidents 'before' would have materialized in the after period had there been no treatment. Similarly, if one is using a comparison group, one predicts by multiplying the count of before accidents by the ratio of after-to-before comparison accidents. The assumption now is that this ratio would be the same for the treated entities and for the entities of the comparison group. The recognition is unifying because what may at first appear to be a multitude of diverse study designs is now seen as the many ways of doing the same thing. The recognition is liberating, because the methods now used for prediction appear rather simple minded. Surely there are other, perhaps better, ways to predict. A mind freed from the fetters of tradition will identify, explore and find improved approaches to prediction.

Since there are many ways to predict, and therefore many possible predictions, there will be many possible estimates of the safety effect of some treatment, all based on the same data. This is a somewhat distressing state of affairs. It acts to diminish one's faith in evaluative research and confidence in the body of knowledge based on it. It is therefore an urgent task to determine which of the many ways to predict are superior.

Without wishing to put bounds on the search for better prediction, it appears that there are four strategic questions that need to be addressed. First, one needs to harness into prediction all that is known. Thus, e.g., if we know how traffic flow has changed from 'before' to 'after', and we know what the relationship between traffic flow and safety is, surely one should predict using this information. Also, if we have many years of accident counts for treated entities, there is no excuse for making use of only the most recent two or three years of data. Second, one needs to recognize that the passage of time influences safety through a variety of causal influences that are not measured or not understood. This too cannot be disregarded in prediction. Third, in observational studies there is always the possibility of a selection bias. If one can say why a treatment has been applied here and not there, there is potential selection bias. Prediction in observational studies must account for this possibility. Fourth, the passage of time, the act of intervening and geographic variation all may influence the probability that a reportable accident is reported. This too must be remembered in prediction.

The issues summarized here and in the previous chapters are the first part of the book. The name given to this part was 'Essentials'. Fortunately it was possible to discuss all that is important in non-mathematical terms. The conclusions reached will guide the rest of this book. The next part is devoted to the adaptation of conventional approaches to the realities of observational studies; the final part will aim to free the task of learning about the safety effect of treatments from conventions that do not suit reality. An attempt will be made to discuss what is important without the encumbrance of formulae. However, since matters that are statistical in essence are at the core of both prediction and estimation, use of mathematics neither can nor should be shunned.

PART II

ADAPTATIONS OF CONVENTIONAL APPROACHES

The first part of this monograph dealt with general issues such as the logic of Before-After studies, definition of safety, accident counts as raw material, and the central role that prediction plays. This second part is an attempt to adapt **conventional approaches** to the realities of **observational studies**. Conventional approaches have been inspired either by the habits of the physics laboratory or the attitudes and methods inherited from the tradition of randomized statistical experiments. Neither heritage fits the circumstances of observational studies. There is no way to arrest change and vary one factor at a time as in the laboratory; nor is it practical to randomly select road sections where, say, the speed limit is to be changed. This is why the methods for extracting information from laboratory data or randomized trials do not necessarily apply to the interpretation of data from observational studies.

At this point the discussion turns quantitative. The aim is to estimate the size of the change in safety and to describe how accurately the size is estimated. This involves mathematical reasoning and notation; neither the use of algebra nor statistics can be avoided. The meaning of symbols will be defined where needed. A glossary is provided to ease the strain on memory. As a rule, derivations and proofs will be at the end of each section. For readers not interested in mathematical deductions, I will try to give the narrative information needed to explain the main lines of reasoning. Principal results will be summarized and illustrated by numerical examples. I am hoping that this will make the main lessons accessible to the wider audience.

The study of conventional approaches is important because much of what has been done, and much of what will be done in the future, fits this mold. In Chapter 6, I provide a unified framework for all Before-After studies. Chapter 7 is devoted to the 'Naive Before-After' study, one where no comparison group is used. The discussion serves to point out the weaknesses of the naive study, which is so frequently found in the professional literature; it also establishes a benchmark for the attainable statistical accuracy. In Chapter 8 I show how changes in traffic flow and changes in similar factors can be accounted for. In Chapter 9 I discuss how to account for other factors by using a comparison group. Chapter 10 is devoted to the question of how to combine results from several entities, sites or studies.

CHAPTER 6

BASIC BUILDING BLOCKS

In Part I of this monograph I have said that to estimate what effect some treatment had on the safety of an entity[1], one has to compare what would have been the safety of the entity in the 'after' period, had treatment not been applied, to what the safety of the treated entity in the 'after' period was. In this, the safety of an entity during a certain period has been defined to be the number of accidents (by kind and consequence) expected to occur on it, usually per unit of time. Thus, safety is an accident frequency measured in accidents per year. Naturally, the effect of the treatment is sought in the change of the expected frequency and severity of target accidents.

6.1 THE FOUR-STEP

The many variants of Before-After studies are merely different ways of doing two tasks:

Task 1. **Predict** what would have been the safety of an entity in the 'after' period, had treatment not been applied and;

Task 2. **Estimate** what the safety of the treated entity in the 'after' period was.

The goal of this section is to furnish a unified framework for the statistical interpretation of all variants of observational Before-After studies. This framework will guide all subsequent discussions. This universal schema will consist of four basic steps. To formulate this four-step requires some notation and a few basic results.

---

[1] The notion of 'entity' is a flexible one. Usually some treatment is applied to several intersections, road sections, drivers, etc. Each can be thought of as an entity. However, a set of intersections, road sections or drivers can also be thought of as one (composite) entity.

Let π be what the expected number of target accidents of a specific entity in an 'after' period[1] would have been had it not been treated; π is what has to be **predicted** and

λ be the expected number of target accidents of the entity in the 'after' period; λ is what has to be **estimated**[2].

The effect of the treatment on safety is judged by comparing λ and π. To compare the two we may be interested in, say,

$\delta \doteq \pi - \lambda$ the reduction in the 'after' period of the expected number of target accidents (by kind or severity), or

$\theta \doteq \lambda/\pi$ the ratio of what safety was with the treatment to what it would have been without the treatment - the 'index of effectiveness.'

When $\theta<1$, the treatment is effective; when $\theta>1$ it is harmful to safety. Also, $100\times(1-\theta)$ is the percent reduction in the expected accident frequency.

To illustrate, suppose that a treatment was implemented at the beginning of 1992. Suppose further that had it not been implemented, one would expect that in the two years 1995 and 1996 there would have been 360.6 target accidents (π=360.6), and that with the treatment in place the expected number in 1995-1996 was 295.2 target accidents (λ=295.2). The reduction in the expected number of accidents in 1995 and 1996, δ(1995-1996)=360.6-295.2=65.4 target accidents in these two years or 32.7 target accidents/year. The index of effectiveness for 1995-1996 θ(1995-1996)=295.2/360.6=0.82. In general, the safety effect of a treatment changes with time. Therefore, the four parameters λ, π, δ and θ are to be thought of as depending on time. This dependence will be made explicit by notation (e.g., as $\lambda(t)$) when necessary.

As indicated by the non-integer numbers in the illustration, λ and π are expected values[3]. Expected values are never known, but can be estimated from observed data. Thus, e.g., if 273 target accidents were recorded in the 'after' period, we would estimate $\hat{\lambda}$=273 even though, if it were

---

[1] The 'after period' is a period of time after the treatment has been implemented. When it ends depends on the circumstances of the study and on whether the treatment effect changes with time.

[2] In principle, we are interested in a vector of values $\lambda_1, \lambda_2, \ldots, \lambda_i, \ldots$, one value for each accident type or accident severity. However, in most cases it is sufficient to think that, at any specific stage of an investigation, the effect of a treatment on one accident kind or severity is being evaluated. Thus, the notation need not be burdened by an additional subscript.

[3] An average that would obtain in a very large number of 'trials'.

possible to repeat the same 'after' period a very large number of times, the average accident count would be 295.2 accidents. Estimates will be designated by a caret above the symbol. Thus, $\hat{\lambda}$ will mean 'estimate of λ'. In the final account, we wish to have the estimates $\hat{\delta}$ and $\hat{\theta}$. Therefore, I have to chart the way leading from data to $\hat{\delta}$ and $\hat{\theta}$. I want to chart the way in such a manner that the final estimates are unbiased, and that their precision is known. This is essential because, if the bias of an estimate cannot be eliminated and the precision of an estimate cannot be described, they can hardly serve as a basis for defensible decisions. The two progressions in Figure 6.1 show that the two basic parameters λ and π feed into functions δ and θ of interest, and that data are first converted into estimates $\hat{\lambda}$ and $\hat{\pi}$ and these, in turn , lead to the end product.

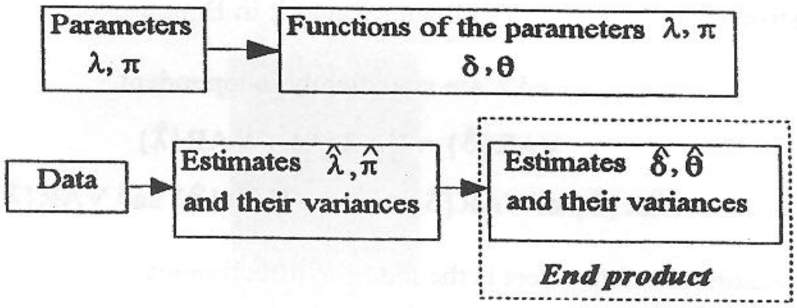


Figure 6.1. From data to end product.

As indicated in Figure 6.1, the estimates $\hat{\lambda}$ and $\hat{\pi}$ are obtained from data. The better the data and the methods of estimation and prediction, the lesser will the differences $\hat{\lambda}-\lambda$ and $\hat{\pi}-\pi$ tend to be. In the earlier illustration we had $\hat{\lambda}=273$ and $\lambda=295.2$. Therefore $\hat{\lambda}-\lambda=-22.2$ target accidents. The usual descriptor of such differences is the variance of the estimate, or its square root - the standard deviation. Thus, e.g., the variance of $\hat{\lambda}$ will be denoted by $VAR\{\hat{\lambda}\}$ and the standard deviation of $\hat{\lambda}$ by $\sigma\{\hat{\lambda}\}$. Estimates $V\hat{A}R\{\hat{\lambda}\}$ of $VAR\{\hat{\lambda}\}$ and $V\hat{A}R\{\hat{\pi}\}$ of $VAR\{\hat{\pi}\}$ are also obtained from data. To illustrate, if we estimate $\hat{\lambda}$ by the 'after' accident count, and if these accident counts are Poisson[1] distributed with a mean λ, then $VAR\{\hat{\lambda}\}=\lambda$. Note that $VAR\{\hat{\lambda}\}$ is an unknown parameter. Since we estimated $\hat{\lambda}=273$ accidents, we may estimate $VAR\{\hat{\lambda}\}$ as $V\hat{A}R\{\hat{\lambda}\}=273$ accidents. The distinction between symbols with and without a caret can be confusing. Unfortunately, if omitted, even worse confusion results. Greek letters and the notations $E\{.\}$ and $VAR\{.\}$ will be used for unknown parameters and the mean and variance of a random variable; Greek letters with a caret and the notations $\hat{E}\{.\}$ and $V\hat{A}R\{.\}$ will be used for estimates obtained from data.

---

[1] By the Poisson distribution, the probability to obtain an accident count L (L is a non-negative integer) is given by $\lambda^L e^{-\lambda}/L!$. In this expression λ is the mean (or expected) accident count.

The difference between the many variants of Before-After studies resides entirely in the methods by which $\hat{\lambda}$ and $\hat{\pi}$ are obtained. These methods, in turn, determine what expressions are to be used for VAR$\{\hat{\lambda}\}$ and VAR$\{\hat{\pi}\}$. This will be discussed in detail in the chapters to follow. In the present chapter I assume that we have the estimates $\hat{\lambda}$, $\hat{\pi}$, $\widehat{\text{VAR}}\{\hat{\lambda}\}$ and $\widehat{\text{VAR}}\{\hat{\pi}\}$, however obtained, and that the task is to use these to construct estimators for δ, θ, VAR$\{\hat{\delta}\}$ and VAR$\{\hat{\theta}\}$. The result will be a common platform for all variants of Before-After studies. The only assumption needed in this chapter is, that the estimates $\hat{\lambda}$ and $\hat{\pi}$ are statistically independent.

1. One measure of the effect of some treatment on safety is δ, the reduction in the expected number of accidents:

$$\delta \doteq \pi - \lambda \qquad \ldots 6.1$$

To obtain an estimate $\hat{\delta}$ of δ, use the estimates $\hat{\lambda}$ and $\hat{\pi}$ in Equation 6.1.

2. Using the assumption that $\hat{\pi}$ and $\hat{\lambda}$ are statistically independent,

$$\text{VAR}\{\hat{\delta}\} = \text{VAR}\{\hat{\pi}\} + \text{VAR}\{\hat{\lambda}\} \qquad \ldots 6.2$$

To obtain an estimate $\widehat{\text{VAR}}\{\hat{\delta}\}$ of VAR$\{\hat{\delta}\}$, use the $\widehat{\text{VAR}}\{\hat{\pi}\}$ and $\widehat{\text{VAR}}\{\hat{\lambda}\}$ in Equation 6.2.

3. The other measure of safety effect is the index of effectiveness θ defined as λ/π. It is common to estimate θ by $\hat{\theta} = \hat{\lambda} / \hat{\pi}$. However, even if $\hat{\lambda}$ and $\hat{\pi}$ are unbiased estimates of λ and π, the ratio $\hat{\lambda}/\hat{\pi}$ is a biased[1] estimate of θ. Though the bias is often small, to remove it is a worthwhile precaution. An approximately unbiased[2] estimator for θ is given by

$$\theta^* = (\lambda/\pi) / [1+\text{VAR}\{\hat{\pi}\}/\pi^2] \qquad \ldots 6.3$$

To obtain an estimate $\hat{\theta}$, use the estimates $\hat{\lambda}$, $\hat{\pi}$ and $\widehat{\text{VAR}}\{\hat{\pi}\}$ in Equation 6.3. The expression in the denominator is a correction factor which is usually only slightly[3] larger than 1.

---

[1] Where the bias comes from can be illustrated as follows. Suppose that a random variable X takes on the values x=0.5 and x=1.5 with equal probability so that E{X}=1. The expected value of 1/X is then (1/0.5+1/1.5)/2=1.333 which is not 1/E{X}=1. This is so because equal deviations of x below and above E{X} exert unequal influence on the ratios 1/x.

[2] Where this comes from is explained in the illustration of Section 6.2.

[3] To illustrate, consider the case when π is estimated by the number of 'before' accidents, and these are assumed to be Poisson distributed. Since for the Poisson distribution the mean equals the variance, the correction factor (the denominator in 6.3) is of the form [1+1/count of 'before' accidents]. With 100 'before' accidents this makes 1.01. Thus, were the correction factor not applied in such a case, θ would be overestimated by about 1%, on the average. With more than 500 or so 'before' accidents the correction factor can be safely neglected.

4. Regarding this correction factor in Equation 6.3 as a constant, the variance of $\hat{\theta}$ can be approximated[1] by:

$$VAR\{\hat{\theta}\} \approx \theta^2[(VAR\{\hat{\lambda}\}/\lambda^2)+(VAR\{\hat{\pi}\}/\pi^2)]/[1+VAR\{\hat{\pi}\}/\pi^2]^2 \quad ...6.4$$

To obtain an estimate $V\hat{A}R\{\hat{\theta}\}$, use the estimates $\hat{\theta}$, $\hat{\lambda}$, $\hat{\pi}$, $V\hat{A}R\{\hat{\lambda}\}$ and $V\hat{A}R\{\hat{\pi}\}$ in Equation 6.4.

Taken together, Equations 6.1 to 6.4 provide the necessary statistical framework for all the methods examined in Part II of this book. Equations 6.3 and 6.4 have been obtained by the method of statistical differentials. Since the method will be used repeatedly in later chapters, it is described more fully in Section 6.2. Included in Section 6.2 is the derivation of Equations 6.3 and 6.4.

Numerical Example 6.1.

To provide a tangible anchor, consider a Naive Before-After study with 173 accidents in the 'before' year and 144 accidents in the 'after' year. In a Naive study one predicts that, had the treatment not been implemented, one would expect in the one-year 'after' period 173 accidents ($\hat{\pi}$=173). For reasons to be explained shortly, the variance of this prediction is estimated to be also 173 (i.e., $V\hat{A}R\{\hat{\pi}\}$=173). Similarly, $\hat{\lambda}$=144 accidents and $V\hat{A}R\{\hat{\lambda}\}$=144 accidents[2]. Using these, $\hat{\delta}$=173-144=29 accidents and $\hat{\theta}=(144/173)/(1+173/173^2)=0.83$.

Also, $V\hat{A}R\{\hat{\delta}\} \approx 173+144=317$ (accidents)^2 and $V\hat{A}R\{\hat{\theta}\}=0.83^2(144/144^2+173/173^2)/(1+173/173^2)^2=0.0087$. That is, we estimate that there has been a reduction of 29 accidents with a standard deviation of $\sqrt{317}$=18 accidents or, alternatively, a 17% reduction (from 100×(1-0.83)) with a standard deviation of 100×$\sqrt{0.0087}$=9%. That is, $\hat{\delta}$=29 accidents, $\hat{\sigma}\{\hat{\delta}\}$=18 accidents, $\hat{\theta}$=0.83, $\hat{\sigma}\{\hat{\theta}\}$=0.09.

From an examination of Equations 6.2 and 6.4 it is apparent that the precision[2] with which δ and θ are estimated depends on the precision with which λ and π are estimated, that is, on VAR{$\hat{\lambda}$ and VAR{$\hat{\pi}$}. Note that VAR{$\hat{\lambda}$} and VAR{$\hat{\pi}$} play similar roles in both equations. That is, the precision of end-product depends roughly in the same manner on the precision with which we

---

[1] See illustration in Section 6.2.

[2] For practical purposes it is convenient to use the square root of the variance - the standard deviation - because its dimension is the same as that of $\hat{\delta}$ and of $\hat{\theta}$. Therefore, while analysis is in terms of variances, standard deviations will be used for final results in numerical examples.

predict 'what would have been . . . ' ($\hat{\pi}$), and the precision with which we estimate "what was . . . " ($\hat{\lambda}$). What VAR{$\hat{\lambda}$} and VAR{$\hat{\pi}$} are, depends on how λ and π are estimated.

It is customary to estimate λ by the count of 'after' accidents. (As has been suggested in the discussion of Figure 5.2, better estimates of λ could be produced.) It is also customary to assume that accident counts obey the Poisson probability law[1]. If so, since the mean and the variance of the Poisson distribution are the same, the count of 'after' accidents estimates λ and also VAR{$\hat{\lambda}$}. To produce a prediction $\hat{\pi}$ is usually more complicated. The complication is due to the fact that we are dealing with an eventuality that did not materialize and therefore could not have been observed. In addition, there are many ways to predict 'what would have been'. Methods of prediction range from assuming that the future would resemble the past, through various time-trend projection, to accounting for changes in traffic or other autonomous factors and compensation for selection bias. Accordingly, the expression for VAR{$\hat{\pi}$} will depend on, and will have to be tailored to, the method of obtaining the prediction $\hat{\pi}$.

The statistical schema can now be described as a sequence of activities. To estimate what effect a treatment had on safety, the statistical analysis can always be thought to consist of the following four basic steps:

FOUR-STEP FOR A SINGLE ENTITY.

STEP 1. **Estimate λ and predict π.** For the time being λ will be estimated from the counts of 'after' accidents. The prediction of π will depend on the method chosen to predict what would have been the expected number of target accidents in the 'after' period had the treatment not been implemented. Several common prediction methods will be examined.

STEP 2. **Estimate VAR{$\hat{\lambda}$} and VAR{$\hat{\pi}$}.** The estimates of VAR{$\hat{\lambda}$} and VAR{$\hat{\pi}$} depend on the method chosen to estimate λ and π. We will assume that the count of accidents is Poisson distributed and therefore VAR{$\hat{\lambda}$}=λ. The method of statistical differentials (see Section 6.2) will be used repeatedly to find an approximate value for VAR{$\hat{\pi}$}.

---

[1] By the Poisson probability law, the probability to have a count of L accidents in a certain period when λ is the expected number of accidents during that period is $e^{-\lambda}\lambda^{L}/L!$.

STEP 3. **Estimate δ and θ using $\hat{\lambda}$ and $\hat{\pi}$ from STEP 1 and $\hat{VAR}\{\hat{\pi}\}$ from STEP 2 in $\delta = \pi - \lambda$ and in $\theta^* \approx (\lambda / \pi)/ [1+VAR\{\hat{\pi}\}/\pi^2]$ (Equations 6.1 and 6.3).**

STEP 4. **Estimate $VAR\{\hat{\delta}\}$ and $VAR\{\hat{\theta}\}$ using $\hat{\lambda}$ and $\hat{\pi}$ from STEP 1 and using $\hat{VAR}\{\hat{\lambda}\}$, $\hat{VAR}\{\hat{\pi}\}$ from STEP 2 in $VAR\{\hat{\delta}\}=VAR\{\hat{\pi}\}+ VAR\{\hat{\lambda}\}$ and in $VAR\{\hat{\theta}\} \approx \theta^2[(VAR\{\hat{\lambda}\}/\lambda^2)+(VAR\{\hat{\pi}\}/\pi^2)]/ [1+VAR\{\hat{\pi}\}/\pi^2]^2$ (Equations 6.2 and 6.4).**

So far I have spoken about estimating the safety effect of some treatment on a single entity. However, usually safety effect is estimated when the same treatment is applied to several entities. Let these be numbered 1, 2, . . . , j, . . . n. Accordingly, in STEP 1 we will obtain the series of estimates $\hat{\lambda}(1), \hat{\lambda}(2), \ldots, \hat{\lambda}(j), \ldots, \hat{\lambda}(n)$ and $\hat{\pi}(1), \hat{\pi}(2), \ldots, \hat{\pi}(j), \ldots, \hat{\pi}(n)$. In STEP 2, the estimated variances of each of these estimates will be obtained. To draw overall conclusions in this case, define

$$\lambda \doteq \Sigma\lambda(j) \text{ and}$$
$$\pi \doteq \Sigma\pi(j) \qquad \ldots 6.5$$

where Σ denotes summation over all 'n' entities. A parallel definition applies to the estimates $\hat{\lambda}$ and $\hat{\pi}$. When the $\hat{\lambda}(j)$ are mutually independent and the $\hat{\pi}(j)$ are also mutually independent, then

$$VAR\{\hat{\lambda}\} = \Sigma VAR\{\hat{\lambda}(j)\} \text{ and}$$
$$VAR\{\hat{\pi}\} = \Sigma VAR\{\hat{\pi}(j)\} \qquad \ldots 6.6$$

The essence of Equations 6.5 and 6.6 is to regard the 'n' separate entities as one 'composite' entity. This has an important implication for STEPS 3 and 4 in which safety effect is estimated. Naturally, the safety effect that is now estimated is for the composite entity. It is therefore in the nature of a 'total effect' when δ is estimated and 'average effect' when θ is estimated. As long as the interest is in the composite entity, STEP 3 and STEP 4 remain unchanged. In this case, $VAR\{\hat{\delta}$ and $VAR\{\hat{\theta}\}$ describe the precision of the total and average effects.

In Section 9.5 and later Chapter 10 I will recognize that different entities may need different comparison groups and that the safety effect of a treatment varies from entity to entity. At that time a slight change in notation will be introduced[1]. Thus, for a composite entity the four-step takes the following form:

[1] To acknowledge that what is being estimated is an average safety effect for a sample of 'n' treated entities that have been pooled into a composite entity, I will replace θ by $\bar{\theta}$ and δ by $\bar{\delta}$. Also, instead of $VAR\{\hat{\delta}\}$ and $VAR\{\hat{\theta}\}$ I will need to use the more cumbersome $VAR\{\hat{\bar{\delta}}\}$ and $VAR\{\hat{\bar{\theta}}\}$.

FOUR-STEP FOR A COMPOSITE ENTITY.

STEP 1.	**For j=1, . . . , n estimate $\lambda(j)$ and $\pi(j)$.**
STEP 2.	**For j=1, . . . , n estimate VAR$\{\hat{\lambda}(j)\}$ and VAR$\{\hat{\pi}(j)\}$.**
COMPOSITE ENTITY STEP.	**Estimate λ, π, VAR$\{\hat{\lambda}\}$, and VAR$\{\hat{\pi}\}$ using Equations 6.5 and 6.6.**
STEP 3.	**Estimate δ and θ.**
STEP 4.	**Estimate VAR$\{\hat{\delta}\}$ and VAR$\{\hat{\theta}\}$.**

At the conclusion of these four steps, the results of an observational Before-After study are described by [$\hat{\delta}$, V$\hat{\text{A}}$R$\{\hat{\delta}\}$] or by [$\hat{\theta}$, V$\hat{\text{A}}$R$\{\hat{\theta}\}$] or by both (see Endnote). These are the product of STEPS 3 and 4. It is all that is needed to form an opinion about the safety effect of a treatment for an entity[1]. The same estimates are also sufficient for comparing the results of several studies and for combining their results when possible. In addition, the same estimates are what is necessary to perform tests of significance.

Because tests of significance are so common in use, a brief explanation of their absence from this monograph is in order. In road safety, the question of practical interest is: "What is the size of the safety effect and how accurately is it known?" It is not clear to me what practical purpose is served here by asking: "Can it be shown that the safety effect is statistically different from 0?" Therefore, when the question is: "what was the effect of treatment X on safety?", to perform a test of significance is not a straightforward response. For a detailed critique of tests of significance in research about the safety effect of treatments, see Hauer (1983a, 1991a).

Equations 6.3 and 6.4 were pulled out of a hat. The task of the next section is to show how they were obtained.

---

1 Another layer of notation and analysis will be added later when the task will be to describe how variable is the safety effect of a treatment when implemented on several entities.

6.2 STATISTICAL DIFFERENTIALS

The purpose of this section is to enable the interested reader to derive various expressions in the text and to indicate to the ambitious reader how to expand the analysis in directions presently unexplored. The method of statistical differentials is a tool to find approximate expressions for the mean, variance and higher 'moments' of functions of random variables. Let Y be a function of the statistically independent random variables $X_1, X_2, \ldots, X_n$. Expanding the function into a Taylor series around the means of $X_1, X_2, \ldots, X_n$ and preserving only the linear and quadratic terms of the series, it can be shown that

$$E\{Y\} \approx Y + [\Sigma_1^n (\partial^2 Y/\partial X_i^2) VAR\{X_i\}]/2$$
$$VAR\{Y\} \approx \Sigma_1^n (\partial Y/\partial X_i)^2 VAR\{X_i\} \qquad \ldots 6.7$$

where Y and the partial derivatives of Y with respect to X_i are evaluated at the means of the X_i. For more detail, see, e.g., Ang and Tang (1975, pp. 198-199) or Benjamin and Cornell (1970, pp.180-186).

The method has already been applied in Section 6.1. The derivations of Equations 6.3 and 6.4 are given here in full to serve as an illustration.

Illustration 1. Derivation of Equation 6.3.

We begin by asking: "what would be $E\{\hat{\theta}\}$ if we estimated θ by $\hat{\theta} = \hat{\lambda}/\hat{\pi}$? To answer, $\hat{\theta}$ will play the role of Y in Equation 6.7, while $\hat{\lambda}$ and $\hat{\pi}$ will be X_1 and X_2. The two partial derivatives in the first equation of 6.7, when evaluated at the means of $\hat{\lambda}$ and $\hat{\pi}$, are 0 and $2\lambda/\pi^3$. Thus, $E\{\hat{\theta}\} \approx \lambda/\pi + [0 \times VAR\{\hat{\lambda}\} + (2\lambda/\pi\ ) VAR\{\hat{\pi}\}]/2 = (\lambda/\pi)[1 + VAR\{\hat{\pi}\}/\pi\ ]$. [2] It appears that if we estimated θ by $\hat{\lambda}/\hat{\pi}$ repeatedly, the average of the estimates would be larger than λ/π by the factor $[1+VAR\{\hat{\pi}\}/\pi^2]$. But, we would like to estimate θ so that the expected value of this estimate be λ/π. This is why, to make the estimator unbiased, one has to divide by $[1+VAR\{\hat{\pi}\}/\pi^2]$ as in Equation 6.3.

Illustration 2: Derivation of Equation 6.4.

Next we are interested in finding an expression for VAR{$\hat{\theta}$}. We estimate θ by $\hat{\theta}=(\hat{\lambda}/\hat{\pi})$/[correction factor] (see STEP 3). This makes $\hat{\theta}$ a function of $\hat{\lambda}$ and $\hat{\pi}$, two statistically independent random variables. Even though the correction factor is also a random variable, it is usually so close to 1 and varies so little, that it can be considered a constant.In the derivation to follow, $\hat{\theta}$ plays the rôle of Y in Equation 6.7 and $\hat{\lambda}$ and $\hat{\pi}$ play the rôle of the X_i.

The partial derivatives of $\hat{\theta}$ with respect to $\hat{\lambda}$ and $\hat{\pi}$ are $(1/\hat{\pi})$/[correction factor] and $(-\hat{\lambda}/(\hat{\pi})^2)$/ [correction factor]. Evaluating these at their means λ and π the partial derivatives become $(1/\pi)$/[correction factor] and $(-\lambda/\pi^2)$/[correction factor] . Using these in Equation 6.7, VAR{$\hat{\theta}$}$\approx$[$(1/\pi)^2$ VAR{$\hat{\lambda}$}+$(-\lambda/(\pi)^2)^2$VAR{$\hat{\pi}$}]/[correction factor]^2. Since $\theta=\lambda/\pi$, this can be rewritten as VAR{$\hat{\theta}$}$\approx\theta^2$[(VAR{$\hat{\lambda}$}/λ^2)+ (VAR{$\hat{\pi}$}/π^2)]/ [correction factor]^2 which is Equation 6.4.

6.3 Chapter Summary

The logical essence of an observational Before-After study is in the juxtaposition of a prediction ($\hat{\pi}$) of what would have been the expected number of target accidents of an entity in the 'after' period, had a treatment not been implemented, with an estimate ($\hat{\lambda}$) of what the expected number of target accidents of the entity was with the treatment in place. The difference 'prediction-estimate' was called $\hat{\delta}$, the ratio 'prediction/estimate' was denoted by $\hat{\theta}$. The statistical task of an observational Before-After study has been defined as that of finding the estimates $\hat{\delta}$, $\hat{\theta}$ and the variances of these.

General expressions for doing so were given and the entire process has been structured into four basic steps. The hope is that the same basic steps can guide analysis, no matter what method is used to predict. At this point the statistical analysis of some common Before-After study designs can begin.

Suggested Readings.

1. On three alternative ways to estimate θ see: Hauer, E (1992a). A note on three estimators of safety effect. *Traffic Engineering and Control*, **33**, No. 6, 388-393.

2. A masterly justification of tests of significance is by Fisher. It begins by: "A lady declares that by tasting a cup of tea made with milk she an discriminate whether the milk or the tea infusion was added first to the cup" in Fisher, R.A. (1971). *The design of experiments*, Hafner Publishing Company, New York. First published in 1935. A jaundiced view of test of significance is presented in: Edwards, A.W. F. (1972). *Likelihood.* Cambridge University Press. Berkson, J. (1942). Tests of significance considered as evidence. *Journal of the American Statistical Association,* **37**, 325-335. Hauer, E. (1983). Reflections on methods of statistical inference in research on the effect of safety countermeasures. *Accident Analysis and Prevention*, **15**, No. 4, 275-285.

3. The method of statistical differentials is described in: Benjamin, J.R. and Cornell, C.A. (1970). *Probability, statistics and decision for civil engineers.* McGraw-Hill, see pages 180-186 and also in Ang, A. H-S and Tang, W.H. (1975). *Probability concepts in engineering, planning and design.* John Wiley and Sons, see pages 198-199.

Endnote[1].

Two measures of the safety effect of a treatment, δ and θ, were introduced in this chapter. The first, being a difference ($\delta=\pi-\lambda$), describes the size of the effect as an increase or reduction in the expected number of accidents. The second, being a ratio ($\theta=\lambda/\pi$), describes the size of the effect as a relative change. Since both measures are made of the same parameters λ and π, whether δ, θ, or both are estimated in a study is a matter of choice. Which measure is chosen should depend on the purpose of the inquiry. If the question is "what good did the treatment do?", or "what good is it likely to do?", one seems to ask about δ. Although θ may be instrumental in answering the latter question by using $\delta=\pi(1-\theta)$, the size of the safety effect is still described by δ. In spite of this, so it seems, in many Before-After studies θ is the measure of choice. It is often expressed as a 'percent reduction'. At some later time, such estimates of θ enter the professional literature as 'Accident Reduction Factors' or 'Accident Modification Factors'. As pointed out by Brundell-Freij, at this point, when θ is used as an accident modification factor for a treatment, the choice of measure (of δ, θ or of some other function of λ and π) acquires added meaning. The implication now is, that if for an entity or a set of entities the θ of a treatment is estimated to be, say, 0.83 then, one should

---

[1] This endnote is based on presently unpublished work by K. Brundell-Freij.

expect a 17% reduction in accidents if the same treatment was implemented on another entity or set of entities. This extension, Brundell-Freij argues, may not be warranted.

To illustrate, consider the situation in Figure 6.2 in which the same treatment has been applied to entities 1 and 2. Assume that we have accurate estimates of λ and π. Thus, e.g., for entity 1, π=1.0 and λ= 0.7 accidents. In this representation, the vertical distance between the diagonal and a point (π, λ) is δ. The slope of the line joining a point to the origin is θ. Consider first the case when the outcomes are points 1 and 2a. If this outcome were typical, one would conclude that the effect of the treatment is to make the expected frequency of 'after' accidents the same for all treated entities. Neither δ nor θ can be used to represent this eventuality. Consider next the case when the outcomes are points 1 and 2b. If this were the typical outcome for the treatment at hand, one would be led to conclude that it reduces the number of accidents by a constant proportion. Therefore, θ (not δ) should describe the effect of such treatments on safety. Finally, were point 1 and 2c typical outcomes, the conclusions would be that the treatment reduces the number of accidents by a fixed amount. If so, δ (not θ) should be used to characterize the safety effect of this treatment.

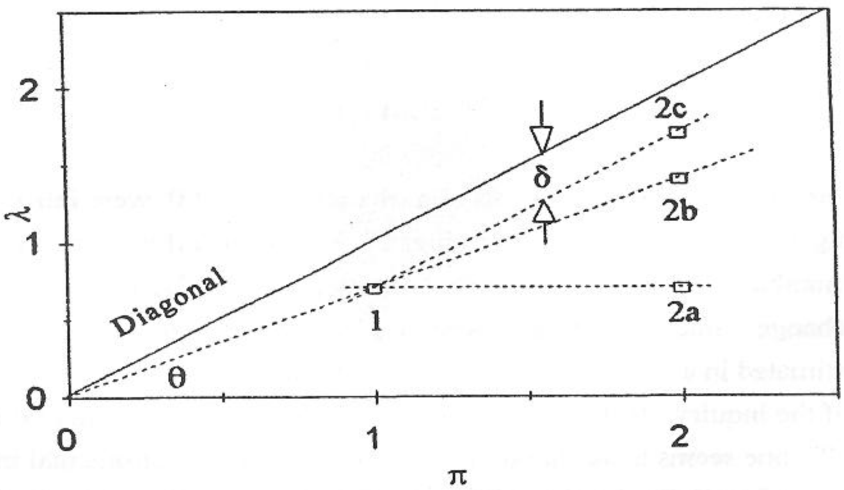


Figure 6.2. Possible treatment effects.

If a treatment is applied to several entities, the outcome for each entity would plot as a point in a Figure such as 6.2. Whether these points will cluster around a line such as 1-2a, 1-2b, or 1-2c is an empirical question. Perhaps the points would not form a line at all, but exhibit a complex functional form. Be it as it may, it seems incorrect to assume a priori that the effect of any treatment is to change the expected number of target accidents by a fixed amount or a by certain fixed proportion.

CHAPTER 7

THE NAIVE BEFORE-AFTER STUDY

In its simplest form, an observational Before-After study consists of comparing the count of the 'before'-period accidents for an entity to its count of 'after'-period accidents. In keeping with the basic logic[1] described in Chapter 2, the count of 'before'-period accidents is used to predict what would have been the expected count of 'after'-period accidents had the treatment not been implemented. This way of predicting reflects a naive and usually unrealistic belief that the passage of time (from the 'before' to the 'after' period) was not associated with changes that affected the safety of the entity under scrutiny. Accordingly, this will be called a 'Naive Before-After' study. In spite of its obvious flaw, the Naive Before-After study deserves thorough discussion. First, because it is a natural starting point for elaboration; one can discuss its merits and shortcomings with clarity. Second, because it is still quite frequently encountered in the professional literature. Third, because it is a useful upper bound, no other study design can reach the statistical precision that is attainable in a Naive Before-After study.

In Chapter 2 I have mentioned the R.I.D.E. program to reduce drinking and driving. Figure 2.1 from Chapter 2 is reproduced here as Figure 7.1. A quick glance is sufficient to show that something has been going on during the 'before' period to reduce the frequency of alcohol-related injuries. Therefore, to assume that all such change would suddenly stop just when R.I.D.E. was implemented is hardly sensible.

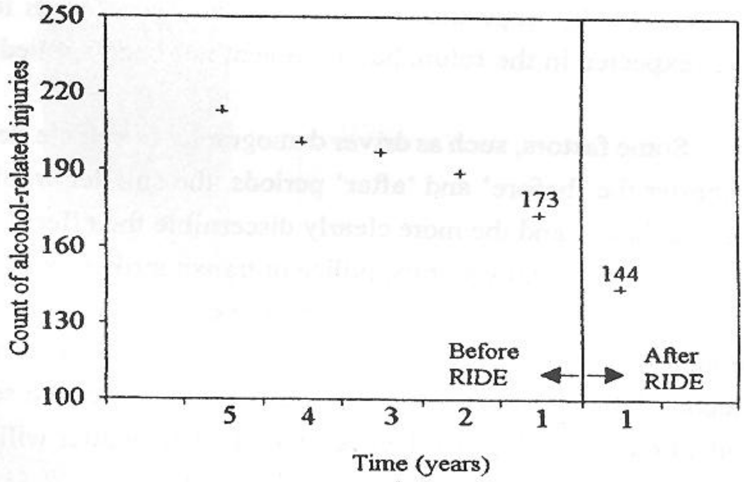


Figure 7.1. Time series of accident counts.

[1] "Compare what would have been the safety of the entity in the 'after' period had treatment not been applied, to what the safety of the treated entity in the 'after' period was."

In this case (and in general), the naive belief that nothing changed from 'before' to 'after' except for the treatment is baseless. Yet this is exactly what is being assumed in a Naive Before-After study.

Five groups of factors make the naive assumption questionable:

a. Traffic, weather, road user behavior, vehicle fleet, and many other factors change autonomously over time. Therefore, the change in safety from 'before' to 'after' surely reflects the effect of change in all these factors, in addition to whatever is due to the treatment.

b. Besides the treatment of interest, various other treatments, programs and treatments may have been implemented at various times during the 'before' or 'after' periods.

c. The count of Property Damage Only accidents is affected by the cost of repairs which change gradually over time. Occasionally the accident count changes suddenly because of adjustments to the reportability limit.

d. The probability of reportable accidents being reported may be changing with time.

e. The entities may have been chosen for treatment because they had unusually many or few accidents in the past. If so, because the past accident history is 'unusual' one can hardly hope that the 'unusual' is a good basis for predicting what would be expected in the future had treatment not been applied.

Some factors, such as driver demography or vehicle fleet, change only gradually. Therefore, the shorter the 'before' and 'after' periods, the smaller would be the influence of such gradually changing factors and the more clearly discernible the effect of the treatment. The same is not true of factors such as snow storms, police or transit strikes, rock concerts, power failures and the like, which may be very different in successive years. Use of short 'before' and 'after' periods does not diminish the worry about the influence of factors that change abruptly. The influence of the 'selection bias' (item e) is the strongest if the entities are selected for treatment because of their accident record during the 'before' period. This matter will be taken up in Chapter 11.

It should therefore be obvious that the Naive Before-After study estimates a mix of what is due to the treatment and what is caused by the sundry influences listed under items 'a' to 'e'. This ought to be made explicit whenever the results of a Naive Before-After study are published, because most readers are not experts in evaluation research. Perhaps the use of the following disclaimer formula is in order:

Disclaimer No.1.

" The noted change in safety reflects not only the effect of . . . (name of treatment) but also the effect of factors such as traffic, weather, vehicle fleet, driver behavior, cost of car repairs, inclinations to report accidents and so on. It is not known what part of the change can be attributed to . . . (name of treatment) and what part is due to the various other influences."

If there is even the slightest suspicion that the decision to treat the entities was influenced by their past accident record, good or bad, and that the same record of past accidents was then used as part of the 'before' data, one should add:

Disclaimer No. 2

" The noted change in safety may be in part due to the spontaneous regression-to-mean and not due to . . . (name of the treatment)."

The statistical analysis in Section 7.1 below can only determine the estimated size of this **mix of effects** (and the statistical accuracy of the estimate of this mix); how much of it is due to the treatment, and how much is due to the other influences cannot be ascertained. That is the main deficiency of a Naive Before-After study. Even if we estimate with statistical precision, what we estimate may not be what we wish to know. To rid the Naive Before-After study of some of its shortcomings will require a series of corrective actions. For these, the analysis of the Naive case is the necessary foundation.

7.1 STATISTICAL ANALYSIS OF THE NAIVE BEFORE-AFTER STUDY

Some treatment has been implemented on entities numbered 1, 2, . . . , j, . . . , n. During the 'before' periods the accident counts were K(1), K(2), . . . , K(n) and during the 'after' periods the accident counts were L(1), L(2), . . . , L(n). To keep notation simple, I will use capital letters to denote both, the name of the random variable and also specific outcomes. The duration of the 'before' and 'after' periods may differ from entity to entity[1]. I define the 'ratio of durations' to be:

[1] Some seem to think it necessary that the 'before' periods and the 'after' periods must be equal for all entities. This leads them to discard useful data for no good reason.

$$r_d(j) \doteq \frac{\textit{Duration of after period for entity } j}{\textit{Duration of before period for entity } j}$$

The estimates $\hat{\lambda}$, $\hat{\pi}$, $\hat{VAR}\{\hat{\lambda}\}$ and $\hat{VAR}\{\hat{\pi}\}$ for a Naive study are given by in Table 7.1. How they were obtained is shown in detail under 'derivations' at the end of this section. These estimates are the subject of STEP 1, STEP 2 and the Composite Entity Step (CES) of the four-step formulated in Chapter 6.

Table 7.1. Estimation when not all $r_d(j)$ are the same.

Estimates of Parameters STEP 1 & CES	Estimates of Variances STEP 2 & CES
$\hat{\lambda}=\Sigma L(j)$ $\hat{\pi}=\Sigma r_d(j)K(j)$	$\hat{VAR}\{\hat{\lambda}\}=\Sigma L(j).$ $\hat{VAR}\{\hat{\pi}\}=\Sigma r_d(j)^2K(j)$

For the common but special case when all r_d's are equal:

Table 7.2. Estimation when all $r_d(j)$ are the same.

Estimates of Parameters STEP 1 & CES	Estimates of Variances STEP 2 & CES
$\hat{\lambda}=\Sigma L(j)$ $\hat{\pi}=r_d\Sigma K(j)$	$\hat{VAR}\{\hat{\lambda}\}=\Sigma L(j).$ $\hat{VAR}\{\hat{\pi}\}=r_d^2\Sigma K(j)$

The next two steps are always as defined in Section 6.1. The expressions are repeated in Table 7.3 for convenience.

Table 7.3. Steps 3 and 4.

$\delta \equiv \pi-\lambda$	... 6.1
$VAR\{\hat{\delta}\} = VAR\{\hat{\pi}\} + VAR\{\hat{\lambda}\}$	... 6.2
$\theta^* = (\lambda/\pi)/[1+VAR\{\hat{\pi}\}/\pi^2]$	... 6.3
$VAR\{\hat{\theta}\} \approx \theta^2[(VAR\{\hat{\lambda}\}/\lambda^2)+(VAR\{\hat{\pi}\}/\pi^2)]/[1+VAR\{\hat{\pi}\}/\pi^2]^2$	... 6.4

Using the estimates $\hat{\lambda}$, $\hat{\pi}$, $\hat{VAR}\{\hat{\lambda}\}$ and $\hat{VAR}\{\hat{\pi}\}$ from Table 7.1 or 7.2, instead of the parameters on the right-hand side of the expressions in Table 7.3, estimates of δ, θ, $VAR\{\hat{\delta}\}$ and $VAR\{\hat{\theta}\}$ are obtained. Since the 'n' entities were pooled into one 'composite entity', the estimates of δ, θ, $VAR\{\hat{\delta}\}$, and $VAR\{\hat{\theta}\}$ pertain to this composite entity. The meaning of these can be explained by the following example. Suppose that two entities were treated. For entity 1 we had a 1-year 'after' period and we estimate $\delta(1)$ to be 6.2 accidents in that year. For entity 2 we had a 3-year 'after' period and estimate δ to be 9.7 accidents in three years. Thus, the estimated expected reduction in accidents for this composite entity is 15.9 in four 'after' years. When the θ for the composite entity is estimated, Equation 6.3 accounts for the difference in the after-period durations automatically. Naturally, both estimates have clear meaning only when the safety effect of the treatment does not change with time. If it does, separate estimates of δ and θ are needed for period 1, period 2, . . . etc. after treatment.

The use of the results in Tables 7.1 and 7.3 will now be illustrated by two examples. In these examples and elsewhere, the results are given in the form of estimates and their standard deviations. In the normal distribution, about two thirds of the probability mass is within ± one standard deviation of the mean, and about 95% of the probability mass is within ± two standard deviations of the mean (see, e.g., Benjamin and Cornell, 1970, p. 141). This is the basis of a useful rule of thumb that allows one to assess the meaning of specific numerical results.

Numerical Example 7.1. Safety effect of R.I.D.E., periods of equal duration.

In one of five police districts in metropolitan Toronto an enforcement program to reduce impaired driving (R.I.D.E.) has been introduced. One year before the program started there were 173 alcohol-related injuries. During the first year that the R.I.D.E. program was in place, there were 144 alcohol-related injuries. We wish to estimate the safety effect of R.I.D.E.. Before computation begins, clarification of what is being computed is in order. None of the four 'beliefs' listed under 'a' to 'd' at the beginning of this chapter can be justified. Therefore the estimates of δ and θ will reflect the influence of a variety of factors, not only the effect of R.I.D.E.. In this case the 'before' and 'after' periods are of equal durations and therefore $r_d=1$. The results are given in Table 7.4 and follow the equations in Tables 7.2 and 7.3, except that the variances are replaced by standard deviations.

Table 7.4. Estimated values.

Estimates of Parameters	Estimates of Standard Deviations
$\hat{\lambda}$=144 a.r.i.*	$\hat{\sigma}\{\hat{\lambda}\}$=12.0 a.r.i.
$\hat{\pi}$=173 a.r.i.	$\hat{\sigma}\{\hat{\pi}\}$=13.2 a.r.i.
$\hat{\delta}$=173-144=29 a.r.i.	$\hat{\sigma}\{\hat{\delta}\}=\sqrt{(173+144)}$=17.8 a.r.i.
$\hat{\theta}$=(144/173)/(1+1/173)=0.83	$\hat{\sigma}\{\hat{\theta}\} \approx (0.83)\sqrt{(1/144+1/173)}/(1+1/173)$=0.09

* alcohol-related injuries

Thus, we estimate that there has been a reduction of 29 in the expected annual number of alcohol-related injuries with a standard deviation of 17.8, or a 17% reduction with a standard deviation of 9%. This **reduction cannot be attributed to R.I.D.E. only** because many other factors changed at the same time. The correction factor 1+1/173=1.006 is in this case negligible (see Equation 6.3).

Numerical Example 7.2. Differing ratios of duration.

At five intersections the accident counts were obtained before and after some treatment during periods of unequal durations. The data are in the first five columns of Table 7.5. The last three columns contain the computations for the first two steps. These follow Table 7.1.

Table 7.5. Data and computations.

Intersection Number	Years Before	Years After	Acc. Before	Acc. After			
j			K(j)	L(j)	$r_d(j)$	$r_d(j)K(j)$	$r_d^2(j)K(j)$
1	3	1	31	7	0.33	10.33	3.44
2	3	1	23	4	0.33	7.67	2.56
3	2	1	7	1	0.50	3.50	1.75
4	2	1	8	5	0.50	4.00	2.00
5	1	1	5	7	1.00	5.00	5.00
Sums				24		30.50	14.75

Thus, $\hat{\lambda}=\Sigma L(j)=24$ accidents in the 'after' year, $\hat{\pi}=\Sigma r_d(j)K(j)=30.5$ accidents in the 'after' year; $\hat{VAR}\{\hat{\lambda}\}=\Sigma L(j)=24$ [accidents]^2 and $\hat{VAR}\{\hat{\pi}\}=\Sigma r_d(j)^2K(j)=14.75$ [accidents]^2. Using now Equations 6.1 to 6.4, $\hat{\delta}=\hat{\pi}-\hat{\lambda}=30.5-24=6.5$ accidents and $\hat{VAR}\{\hat{\delta}\}=\hat{VAR}\{\hat{\pi}\}+\hat{VAR}\{\hat{\lambda}\}=14.75+24=38.75$ [accidents]^2. Also $\hat{\theta}=(\hat{\lambda}/\hat{\pi})/[1+\hat{VAR}\{\hat{\pi}\}/\hat{\pi}^2]=(24/30.5)/(1+14.75/30.5^2)=0.775$ and $\hat{VAR}\{\hat{\theta}\}\approx\hat{\theta}^2[(\hat{VAR}\{\hat{\lambda}\}/\hat{\lambda}^2)+(\hat{VAR}\{\hat{\pi}\}/\hat{\pi}^2)]/[1+\hat{VAR}\{\hat{\pi}\}/\hat{\pi}^2]^2=0.775^2[1/24+14.75/30.5^2]/(1+14.75/30.5^2)=0.034$. This is summarized in the standard form in Table 7.6.

Table 7.6. Estimated values

Estimates of Parameters	Estimates of Standard Deviations
$\hat{\lambda}=24$ accidents	$\hat{\sigma}\{\hat{\lambda}\}=4.9$ accidents
$\hat{\pi}=30.5$ accidents	$\hat{\sigma}\{\hat{\pi}\}=3.8$ accidents
$\hat{\delta}=6.5$ accidents	$\hat{\sigma}\{\hat{\delta}\}=6.2$ accidents
$\hat{\theta}=0.775$	$\hat{\sigma}\{\hat{\theta}\}\approx0.183$

Derivations.

STEP 1 in Table 7.1. The estimates of λ and π.

Here the task is to show where the expressions for $\hat{\lambda}$ and $\hat{\pi}$ come from. Recall that λ denotes the number of accidents expected to occur in the 'after' periods on the composite entity made up of entities 1, 2, ..., n. On this composite entity the count of accidents in the 'after' period is $\Sigma L(j)$. The Σ sign indicates a summation from 1 to n. As usual, this sum is used to estimate λ.

Also recall that π is the notation for the 'prediction'; the number of accidents that would be expected to occur in the 'after' period had the treatment not been implemented. The essence of the Naive method is in the assertion[1] that had the treatment not been implemented, the safety of the 'after' period would be the same as the safety of the 'before' period of equal duration. Let $\kappa(j)$ denote the expected number of accidents in the 'before' period for entity j. If the ratio of the 'after' to 'before' periods for this entity is $r_d(j)$, by the assertion of the Naive method, one would predict that $\pi(j)=r_d(j)\kappa(j)$. Equation 6.5 defines $\pi \doteq \Sigma\pi(j)$. If so, $\pi=\Sigma r_d(j)\kappa(j)$. It is customary [2]to use the count of 'before' accidents K(j) to estimate $\kappa(j)$. Therefore, $\hat{\pi}=\Sigma r_d(j)K(j)$. In sum, as listed in Table 7.1, we estimate

$$\hat{\lambda}=\Sigma L(j)$$
$$\hat{\pi}=\Sigma r_d(j)K(j) \qquad \ldots 7.1$$

In many Naive Before-After studies the ratio of durations is the same for all treated entities. Let r_d denote this common ratio. In this special case, $\hat{\pi}=r_d\Sigma K(j)$.

STEP 2 in Table 7.1. The estimates of VAR{$\hat{\lambda}$} and VAR{$\hat{\pi}$}.

To obtain an expression for VAR{$\hat{\lambda}$} use is made of the first Equation 7.1. On this basis we can write $VAR\{\hat{\lambda}\}=VAR\{\Sigma L(j)\}$. Because accident counts are statistically independent, the variance of a sum is the sum of the variances. That is, $VAR\{\Sigma L(j)\}=\Sigma VAR\{L(j)\}$. Since the accident counts are assumed to be Poisson distributed, $VAR\{L(j)\}=\lambda(j)$. If so, $VAR\{\hat{\lambda}\}=\Sigma\lambda(j)$. As earlier, $\Sigma\lambda(j)$ is estimated by $\Sigma L(j)$.

---

[1] This is in fact projection 1 in Figure 5.1 when a one-year 'before' period is used. That the assertion is quite nonsensical in the context of this figure is self-evident. It is therefore quite amazing to recognize that such a questionable assertion is so commonly accepted as a legitimate way to proceed.

[2] A different estimator will be introduced in Part III.

To obtain an expression for VAR{$\hat{\pi}$} use is made of the second Equation 7.1 from which it follows that VAR{$\hat{\pi}$}=VAR{$\Sigma r_d(j)K(j)$}. Again, since the accident counts are statistically independent, VAR{$\Sigma r_d(j)K(j)$}=ΣVAR{$r_d(j)K(j)$}. Recall that the ratio of durations is a constant. Therefore, VAR{$r_d(j)K(j)$}=$[r_d(j)]^2$VAR{K(j)}. So that, VAR{$\hat{\pi}$} = $\Sigma r_d(j)^2$VAR{K(j)}. Using now the assumption that accident counts are Poisson distributed, VAR{K(j)} is estimated by K(j). If so, we may estimate VAR{$\hat{\pi}$} by $\Sigma r_d(j)^2K(j)$. In sum,

$$\widehat{VAR}\{\hat{\lambda}\}=\Sigma L(j).$$
$$\widehat{VAR}\{\hat{\pi}\}=\Sigma r_d(j)^2K(j) \qquad \ldots 7.2$$

In the special case when all $r_d(j)$'s are the same, $\widehat{VAR}\{\hat{\pi}\}=r_d^2\Sigma K(j)$.

7.2 SEPARATING THE WHEAT FROM THE CHAFF

In Numerical Example 7.1 the conclusion was, that following the implementation of R.I.D.E. the expected annual number of alcohol-related injuries has decreased by 17±9%[1]. But 'following R.I.D.E.' does not mean 'because of R.I.D.E..' The 17% reduction is not only the result of R.I.D.E., but also the consequence of change from the 'before' to the 'after' period in many other factors that influence safety. All this has already been said at the beginning of this chapter. To anchor the correct image in the mind, it may be useful to put the influence of the treatment on one axis, and the influence of all other factors on the other axis in the same graph.

In Figure 7.2 the effect of R.I.D.E. is measured on the horizontal axis and the influence of all other factors on the vertical axis. The solid line at 45° represents all points for which the **sum** of these two effects is constant, a 17% reduction. A Naive Before-After study allows only the estimation of this sum. Thus, all that can be said at this time is that the effect of R.I.D.E. is estimated to be **some point on this line**. This covers all possible effects of R.I.D.E. and amounts to knowing nothing about its effect on safety. Were it known, e.g., that no factors other than R.I.D.E. have changed from the 'before' to the 'after' period, we would estimate the influence of R.I.D.E. to be a 17% reduction (Point A). However, the 'other' factors (traffic, weather, economic activity, cost of liquor etc.) are always present and change in time. Were it known, e.g., that the influence of the 'other' factors caused a reduction of, say, 20% in the expected number of alcohol-related injuries, R.I.D.E. would be deemed to have caused a 3% increase in alcohol-related injuries (Point B). Alternatively, if the influence of all factors other than R.I.D.E. was to increase the expected number of injuries by, say, 5%, R.I.D.E. would be estimated to have caused a 22%

---

[1] In this notation the ± sign preceeds one standard deviation.

reduction in injuries (Point C). It follows that, as long as the influence of the 'other' factors cannot be determined, the influence of R.I.D.E. will remain unknown. The band of ±9% surrounding the heavy line does not describe the uncertainty with which influence of R.I.D.E. is known. It describes the uncertainty with which the joint influence of R.I.D.E. and all other factors is known.

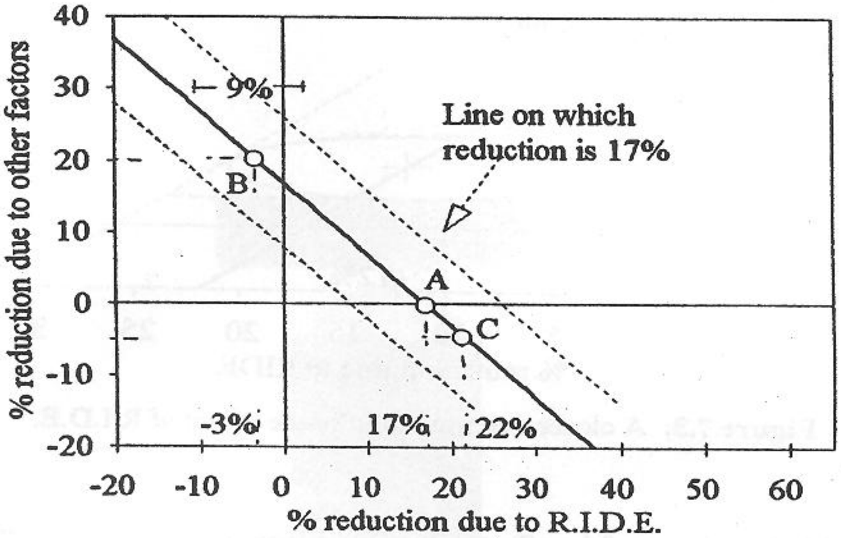


Figure 7.2. Separating the effect of R.I.D.E. from other factors.

Yet, it is entirely feasible to speculate about the effect of factors other than R.I.D.E.. Therefore one can be more specific about the influence of R.I.D.E. than to say: "it is somewhere on the heavy line." To reduce the range of possible values for the influence of R.I.D.E. one has to estimate the effect of factors other than R.I.D.E.. One way of doing so is from historical data. In Figure 5.1 I have shown the count of alcohol-related injuries during five years before R.I.D.E.. These are listed in Table 7.7. The average reduction in injury accidents in four year-pairs before R.I.D.E. was 5.04% with a sample standard deviation of 2.7%. If past reductions due to factors other than R.I.D.E. are an indication of their influence during the 'before' and 'after' treatment years, this can be represented as in Figure 7.3.

Table 7.7. Count of injuries before R.I.D.E.

Year	Count of Injuries	% Reduction
5	213	
4	201	5.63
3	197	1.99
2	189	4.06
1	173	8.47
Average % reduction		5.04
Sample standard deviation		2.73

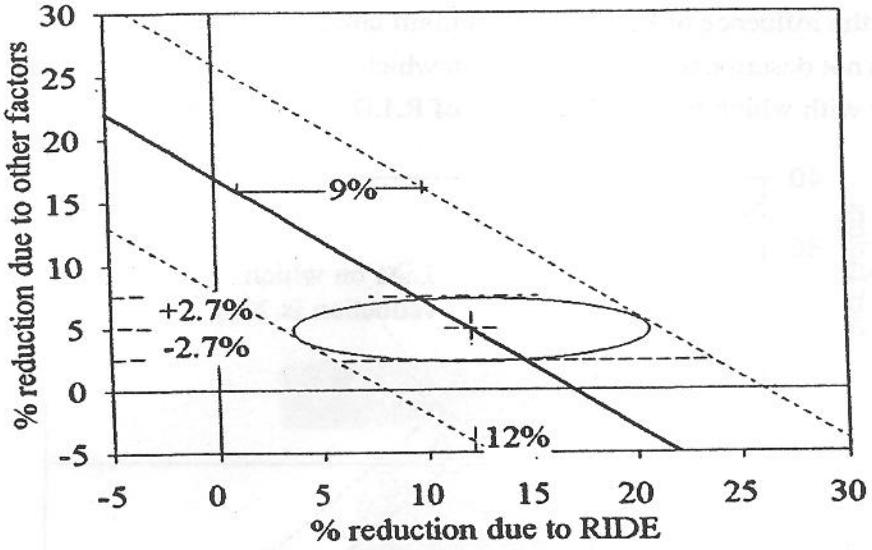


Figure 7.3. A closer determination of the effect of R.I.D.E.

The most likely estimate of the effect of R.I.D.E. is now about a 12% (=17%-5%) reduction in alcohol-related injuries. This point is at the center of the ellipse inscribed into the ±9% band and the ±2.7% band. This ellipse is a rough visual depiction of the uncertainty surrounding the estimated 12% reduction due to R.I.D.E..

A better analysis of this problem is possible . However, the purpose here was to show and to emphasize that, in order to say anything about the effect of a treatment, one has to have an idea what the effect of the 'other' factors was. Otherwise the effect of the treatment is simply **indeterminate and remains unknown.** A Naive Before-After study does not provide any means for determining what was the effect of the 'other' factors. Therefore, unless one is willing to believe that the world stood largely still for the duration of the 'before' and 'after' periods, a Naive study is uninterpretable.

7.3 STUDY DESIGN CONSIDERATIONS

Convention has it that before a study is undertaken, it is designed. This may be true of randomized experiments in which one can buy more laboratory rats for drug testing or recruit more subjects for a clinical trial. Observational studies are opportunistic endeavors. The researcher follows the doer; the material for the study is often limited to what data one can find; the activity preceding data analysis is more akin to scavenging than to deliberate experimental design. However, while random assignment of entities to either treatment or control is not possible, there is often a degree of control over what kind of information one acquires and how much of the available data one elects to use.

In this chapter focus is on the Naive Before-After study, the most Spartan of all observational studies. The only choice that can perhaps be made here is that of the number of accidents. So, before a Naive study is undertaken, one might want to know how many accidents are needed to estimate the change in safety[1] with satisfactory precision. Two decisions determine the number of accidents that will form the database of the study: the **number of entities** (sites, drivers, road sections, etc.) for which accidents are to be counted, and the **duration** of the 'before' and 'after' periods. In the 'derivations' at the end of this section I show that:

$$\Sigma\kappa(j)=(\theta/r_d+\theta^2)/\sigma^2\{\hat{\theta}\} \qquad \ldots 7.3$$

As always, $\Sigma\kappa(j)$ is the sum over all treated entities of the expected number of 'before' accidents, θ the index of effectiveness (Section 6.1), and r_d the ratio of 'after' to 'before' period durations (Section 7.1).

Equation 7.3 can be used to guide deliberations about the number of accidents needed. To appreciate its message, consider the simple case when the 'before' and 'after' periods are of equal duration (r_d=1) and when the effect of the treatment is moderate so that θ is approximately 1. If so,

$$\Sigma\kappa(j)\approx 2/\sigma^2\{\hat{\theta}\}$$

Recall that by the rule of thumb, about 65% of the probability mass is usually within 1σ of the mean, 95% within 2σ, and 99.5% within 3σ. Thus, there is a one-in-twenty chance that the estimate of a parameter ± two standard deviations is not covering the true value of the parameter. With ± three standard deviations the chance is about 1/200. If we wish $\sigma\{\hat{\theta}\}$to be, say, 0.1, then, about 2/0.01=200 'before' accidents are needed. Thus, with 200 'before' accidents the chance of getting $\hat{\theta}$ to be within 0.1 of its true value (θ) is about 65%. If a higher precision is desired, say, if $\sigma\{\hat{\theta}\}$ is to be 0.01, about 2/0.0001=20,000 'before' accidents are needed. There is a message in these simple results. One can often devise a study with about 200 accidents in the 'before' period. In this case the standard deviation of the estimate of θ is about 0.1. With 200 'before' accidents the effect of treatments which change safety by about 10% or less will be difficult to detect because the effect that is sought is approximately the size of the standard deviation of its estimate. The following rule of thumb applies:

---

[1] Recall that in a Naive study the change in safety due to the treatment is mixed in with that due to all other factors.

> *Rule of thumb:*
> ***The standard deviation of the estimate has to be 2-3 times smaller than the effect which one expects to detect.***

Thus, to detect a 10% change in the expected number of target accidents, $\sigma\{\hat{\theta}\}$ has to be less than 0.05 and preferably below 0.03 or so. This requires at least $2/0.05^2=800$ and preferably $2/0.03^2 \approx 2200$ accidents. It follows that, if the treatment is expected to change safety only by a few percentage points, to detect such a change, the number of 'before' accidents must be in the ten-thousands. Since this size is seldom available in practice, one must conclude that:

> ***to reliably detect a change in safety of only a few percentage points requires such a large number of accidents, that the conduct of such a study is rarely practical.***

One can reduce the required number of 'before' accidents by increasing the duration of the 'after' period, thereby increasing r_d. If, e.g., the 'after' period is twice the duration of the 'before' period, $r_d=2$. To have $\sigma\{\hat{\theta}\}=0.1$ in this case one will need 150 (not 200) 'before' accidents. It is plain that little is gained in precision by making the 'after' period longer. However, extending the duration of the 'before' and 'after' periods strains ever more the five 'beliefs' listed early in this chapter. The estimate is less likely to reflect the effect of the treatment and more likely the effect of a variety of unaccounted-for factors that change with the passage of time. Figures 7.4 a, b and c may be used to determine the number of 'before' accidents required to estimate selected values of θ with a given standard deviation.

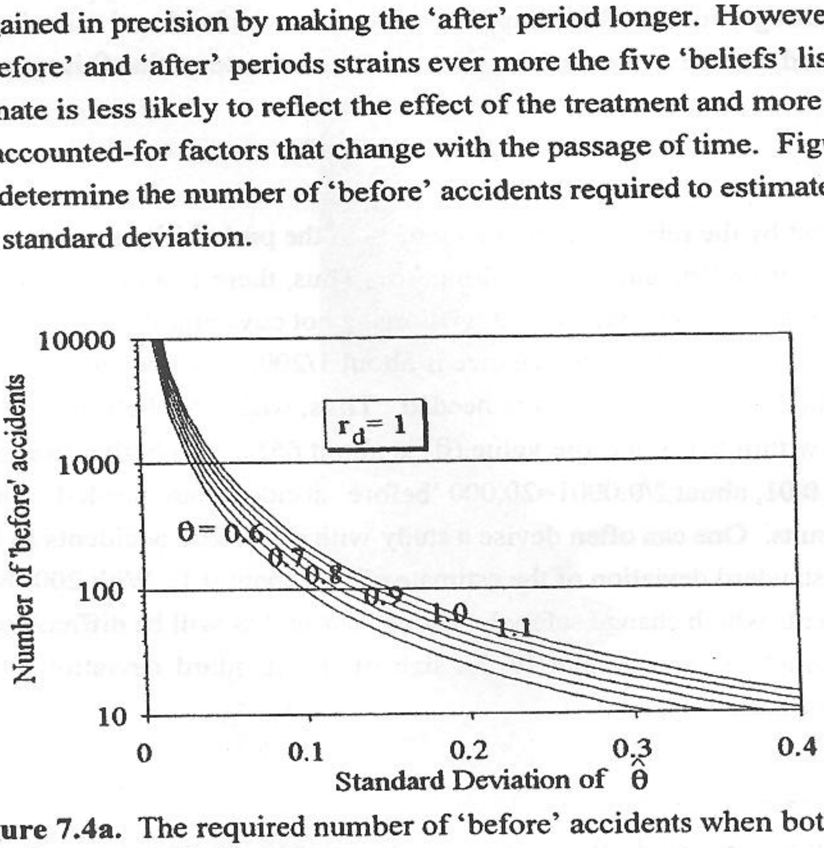


Figure 7.4a. The required number of 'before' accidents when both periods are same in duration.

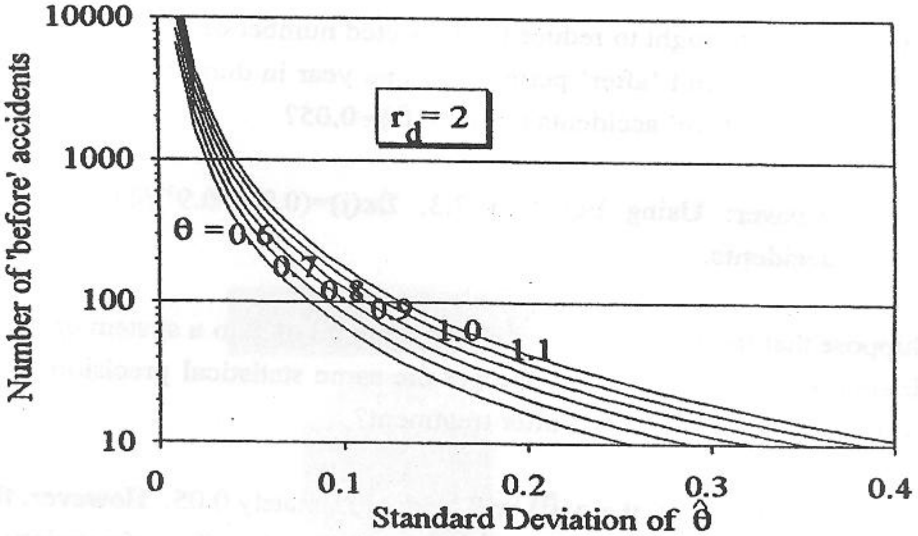


Figure 7.4b. The required number of 'before' accidents when the 'after' period is twice as long as the 'before' period.

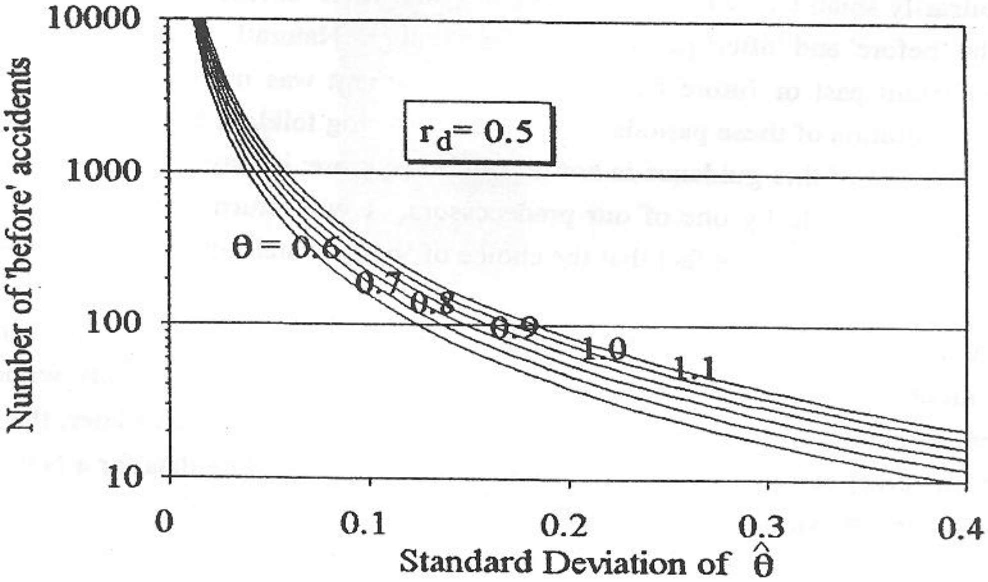


Figure 7.4c. The required number of 'before' accidents needed when the 'before' period is twice as long as the 'after' period.

Numerical Example 7.3. How many accidents are needed?

A treatment is thought to reduce the expected number of accidents by 10% (i.e., θ=0.9). If the 'before' and 'after' periods are one year in duration, what must be the (expected) count of 'before' accidents to get $\sigma\{\hat{\theta}\}\approx0.05$?

> Answer: Using Equation 7.3, $\Sigma\kappa(j)=(0.9/1+0.9^2)/0.0025\approx700$ accidents.

Suppose that the treatment can be administered only to a system on which some 175 accidents occur per year. Can one get the same statistical precision by counting accidents for four years before and after treatment?

> Answer: Yes, the $\sigma\{\hat{\theta}\}$ will be approximately 0.05. However, the estimate of θ will reflect to a larger extent the effect of a variety of factors which will change in the course of the 8 years.

Numerical Example 7.3 underscores the fact that the larger the number of accidents in the 'before' and 'after' periods, the larger is the statistical precision. In speculation, $\sigma\{\hat{\theta}\}$ could be made arbitrarily small by increasing the 'before' and 'after' accident counts $\Sigma L(j)$ and $\Sigma K(j)$ by making the 'before' and 'after' periods longer and longer. Naturally, one hesitates to use data that is in the distant past or future from when the treatment was implemented. What then is an appropriate duration of these periods? Traffic engineering folklore has it, that 3-year periods are best. The source of this guidance is not traceable anymore; it may have its origin in a common sense judgement made by one of our predecessors. I will return to this issue in Part III. The following case illustrates the fact that the choice of 'before' and 'after' period durations matters.

A state highway agency had a program to let contracts for highway resurfacing quickly and without much fuss. A new blacktop was laid on the existing surface usually without changes to alignment, superelevation, shoulders or roadside hazards. Several years later, the question arose whether this practice has caused some deterioration in safety. The data for a Naive Before-After comparison are shown in Figure 7.5.

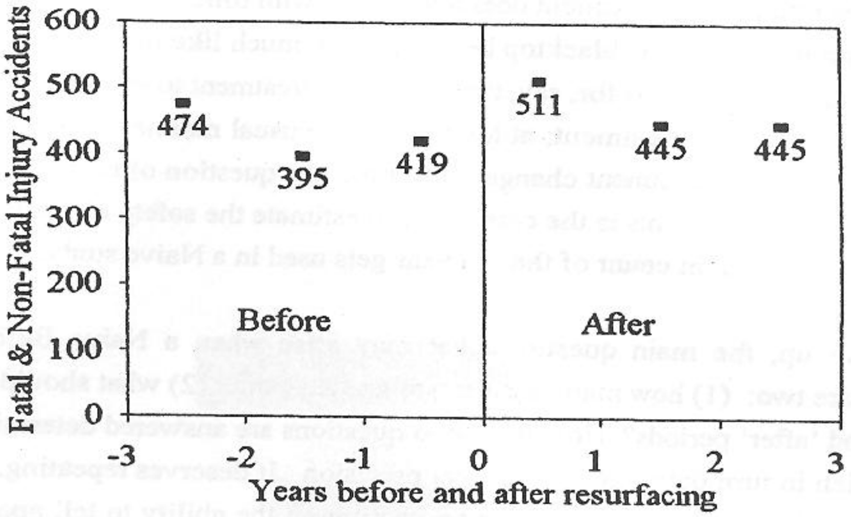


Figure 7.5. Accident counts 'before' and 'after' resurfacing.

Recall that within the confines of a Naive study it is assumed that the count 'before' is a good estimate of π - the expected count of 'after' accidents had no treatment taken place. Were a 1-year 'before resurfacing' period used, the estimate of π would be 419 accidents; with a 2-year 'before resurfacing' period the estimate would be (395+ 419)/2=407 accidents; with a 3-year 'before resurfacing' period it would be 429.3 accidents. Naturally, the conclusion of the study will depend on the choice of the 'before' period duration.

If one could be certain that nothing has changed from year to year, one should take as long a 'before' period as possible. But, an assumption of 'no change' is never entirely correct. In our case, e.g., the 474 count looks suspiciously high. Were some road sections selected for resurfacing because three years earlier they had unusually many accidents? Was there a bad winter in that year? Does the gain in statistical precision from including the 474 accidents in the 'before' count outweigh the possible error of this count being 'unrepresentative'? Clearcut guidance on what is the best choice of 'before' duration is not forthcoming. However, the following three considerations can guide our deliberations.

First, for a Naive study to be at all legitimate, there must be no trend in the 'before' accident counts, and there must be no reasons why the 'before' period years may differ from the 'after' period to which the prediction applies. Second, the more such trend-free and compatible years comprise the 'before' period, the better. Third, as is evident from Figures 7.4, the marginal gain in estimation precision rapidly diminishes with the number of accidents. Thus, one may gain much by using two years instead of one, but it may not be worthwhile to extend the 'before' period from, say, three to four years.

Similar arguments and considerations apply to the choice of the 'after' period duration provided that the effect of the treatment does not change with time. For resurfacing, this is clearly not true. After a while the new blacktop begins to look much like the one it replaced. If so, one cannot use the count of accidents for, say, two years after treatment to estimate the expected number of accidents one year after treatment; at least not in the usual manner of the Naive Before-After study. If the effect of the treatment changes with time, the question of how long the 'after' period should be does not arise. If this is the case, then, to estimate the safety effect in the i-th year after treatment only the accident count of the i-th year gets used in a Naive study.

To sum up, the main questions that may arise when a Naive Before-After study is contemplated are two: (1) how many entities are needed?, and (2) what should be the duration of the 'before' and 'after' periods? How these two questions are answered determines the number of accidents which in turn determines statistical precision. It deserves repeating, that the statistical precision attainable in a Naive study has no bearing on the ability to tell apart the effect of the treatment from the effect of the various other factors that affect safety.

Roughly speaking, there is a one-in-twenty chance that the estimate of a parameter ± two standard deviations is not covering the true value of the parameter. With ± three standard deviations the chance is about 1/200. For this reason one should aim to attain a standard deviation of the estimate that is about a half or a third of the change in safety that we wish to estimate. This rule of thumb leads to the important conclusion that one can seldom amass sufficient accidents to detect safety changes of only a few percentage points.

The longer the 'before' period the better, provided that there is no time trend to safety. If the effect of the treatment changes with time, the question of the 'after' period duration does not arise. Otherwise, in the absence of any time trend, the longer it is the better.

Derivation.

Equation 7.3.

In STEP 4 of Chapter 6 (Equation 6.4) we noted that:

$$\mathrm{VAR}\{\hat{\theta}\} \approx \theta^2[(\mathrm{VAR}\{\hat{\lambda}\}/\lambda^2)+(\mathrm{VAR}\{\hat{\pi}\}/\pi^2)]/[1+\mathrm{VAR}\{\hat{\pi}\}/\pi^2]^2$$

The squared correction factor $[1+\mathrm{VAR}\{\hat{\pi}\}/\pi^2]^2$ is usually close to 1 and, in the present context, can be safely neglected. In addition most studies can be designed assuming that all entities have the same r_d. In Section 7.1 we used the fact that when the accident counts are Poisson distributed and $\hat{\lambda}=\Sigma L(j)$, then $\mathrm{VAR}\{\hat{\lambda}\}=\lambda$. We also noted that if without treatment the safety of the treated entities would not have changed from 'before' to 'after', then, $\pi = r_d \Sigma\kappa(j)$ and $VAR\{\hat{\pi}\} = r_d^2 \Sigma\kappa(j)$. Therefore,

$$VAR\{\hat{\theta}\} \approx \theta^2[1/\lambda + 1/\Sigma\kappa(j)]$$

Using now the identity $\lambda = \theta\pi = \theta r_d \Sigma\kappa(j)$ and rearranging terms,

$$\Sigma\kappa(j) = (\theta/r_d + \theta^2)/VAR\{\hat{\theta}\} \qquad \ldots 7.3$$

7.4 SIGNAL HEADS AND INTERGREEN[1] TIMES - ON READING AND LEARNING

Sections 7.1 and 7.2 were about the analysis of a Naive Before-After study and Section 7.3 dealt with the design of such studies. However, the question before us is not only how to analyze and design a (Naive) Before-After study. Perhaps more important is the task of interpreting reports of work done by others. In this section I will attempt to interpret what is 'on the lines' of a published report and add commentary for what may be 'between the lines'.

Post-mounted signal heads were replaced by mast-mounted signal heads at six intersections. At the same time, a one second 'all-red' phase was added to the intergreen interval (ITE Journal, Vol. 61, No. 10, October 1991). One of the six intersections was excluded from the study because the results there were "considered to be chance fluctuations."

Comment. In the analysis of real data, it is often necessary to decide whether an observation is an 'outlier' and therefore should be removed from the statistical analysis. It is true, however, that by judicious removal of inconvenient data points, one can color the conclusions in desired hues. This is why, the designation of a data point as an 'outlier' must be explicitly justified. Accident counts are, by their very nature, subject to chance fluctuations. This is why in Section 3.2 safety was defined as an expected value. Therefore, if the accident count at one intersection is not similar to that at the remaining five due to chance fluctuations, this is not a sufficient reason for excluding that intersection from further consideration. In principle, the attempt must be made to brand as outliers only those data points which are **different for a reason**, not due to a chance fluctuation. In the present case 'different for a reason' may mean that intersection has a unique geometry, different

[1] 'Intergreen' is the time between when the signal aspect is green for one set of intersection approaches and the other set of approaches which receives green next. 'All-red' is that part of the intergreen when drivers on all approaches see the red signal aspect.

approach speeds, peculiar signal timing, and the like. But, if such reasons are not evident, removal from analysis is not easily justified.

The 'before' and 'after' periods were 12 months in duration. At the five sites included in the study, the following data were obtained:

Table 7.8. Accident counts at five sites.

	Right Angle	Rear End	Left Turn	Other	Total
Before	65	37	37	22	161
After	24	30	50	16	120

To illumine some issues at hand we will consider the right-angle and left-turn accidents. For right-angle accidents the statistical four-step yields:

Table 7.9. Tableau of estimates for right-angle accidents.

Estimates of Parameters	Estimates of Standard Deviations
$\hat{\lambda}$=24 accidents	$\hat{\sigma}\{\hat{\lambda}\}$=4.9 accidents
$\hat{\pi}$=65 accidents	$\hat{\sigma}\{\hat{\pi}\}$=8.1 accidents
$\hat{\delta}$=65-24=41 accidents	$\hat{\sigma}\{\hat{\delta}\}=\sqrt{(65+24)}$=9.4 accidents
$\hat{\theta}$=(24/65)/(1+1/65)=0.36	$\hat{\sigma}\{\hat{\theta}\}\approx(0.36)\sqrt{(1/24+1/65)}/(1+1/65)$=0.09

Evidently, at these five intersections the number of right-angle accidents has diminished. This can be expressed either as a reduction of 41 ± 9.4 accidents in the 'after' year or as a reduction of 64% ± 9%. The uncertainty surrounding both estimates is relatively small. Taking two standard deviations in each direction, the reduction is likely to be in the 22 - 60 accident range or the 46% - 82% reduction range.

Comment. The direction of this finding is in line with what others have found to be the effect of extending the intergreen period by an all-red interval. Also, the moving of signal heads from posts to masts, may contribute to drivers noticing the signal display earlier and help to bring about such a reduction. Of course, it is impossible to say how much is due to the addition of the all-red phase and how much is due to the better placement of signal heads. The author states that the 'before' and 'after' periods were "reasonably comparable because traffic volumes remained fairly constant and no other changes were implemented." This statement does not offer sufficient assurance. There remain unaccounted-for changes in precipitation, enforcement, reportability limits, and all the other factors listed at the beginning of Chapter 7. The reader may want to supplement the findings in her or his own mind by 'Disclaimer No. 1' also stated at the beginning of this chapter. The change noted is due to the two treatments and also due to changes in other important factors. How much is due to which cause is not known.

The author writes that: "This measure results in large safety benefits when it is implemented at signalized intersections . . . experiencing a large number of right-angle collisions prior to the installation of mast-mounted signals." He also writes that one intersection has been excluded from the analysis partly because "the number of right-angle accidents susceptible for correction . . . were very few in the 'before' period." An added statement is that: "Intersections experiencing a large number of right angle accidents prior to . . . showed the greatest reduction in number of accidents."

Comment. These statements should raise a red flag. Have these sites been selected for improvement because they had experienced many right-angle collisions? If the treated intersections had an unusually high number of right-angle accidents in the 'before' period, the entire statistical analysis in Table 7.9 is baseless. Recall that the foundation of the Naive study is the belief that the 'before' accident count can be used to predict what would have been the expected number of accidents for the after period had there been no treatment. But if the 'before' accident count is unusually large (as seems to be the case) this belief is void. One does not predict what is usually expected to occur using unusual occurrences for that purpose. By including in the study intersections experiencing many right-angle collisions and excluding from the study those where such accidents were few, one is virtually guaranteed to see an illusory reduction of the kind noticed (where there were most accidents before treatment, there the illusory effect was largest). This is the familiar regression-to-the-mean bias. This question will be addressed later. Suffice to say here, that if sites were selected to be treated because of an unusually large number of accidents in the 'before' period, the statistical machinery for the analysis of Naive Before-After studies as developed in Section 7.1 is not applicable. Therefore, the reader may want to temper the author's conclusions with 'Disclaimer No. 2'.

The standard tableau for left-turn accidents is given below.

Table 7.10. Tableau of estimates for left-turn accidents.

Estimates of Parameters	Estimates of Standard Deviations
$\hat{\lambda}$=50 accidents	$\hat{\sigma}\{\hat{\lambda}\}$=7.1 accidents
$\hat{\pi}$=37 accidents	$\hat{\sigma}\{\hat{\pi}\}$=6.1 accidents
$\hat{\delta}$=37-50=-13 accidents	$\hat{\sigma}\{\hat{\delta}\}=\sqrt{(37+50)}$=9.3 accidents
$\hat{\theta}$=(50/37)/(1+1/37)=1.32	$\hat{\sigma}\{\hat{\theta}\}\approx(1.32)\sqrt{(1/50+1/37)}/(1+1/37)$=0.28

The author says that: "Left-turn collisions increased by 35%." From the tableau, there was a change of 32% ± 28%.

> *Comment:* The difference between 35% and 32% is due to the 'correction factor' the value of which here is 1/(1+1/37). In Numerical Example 7.1, the correction factor removing the bias from $\hat{\theta}$ proved entirely unimportant. In the present case it alters the estimated value by a few percentage points because the counts of 'before' accidents are relatively small (65 in Table 7.9 and 37 in Table 7.10).

The author finds that the reductions in right-angle and total accidents were statistically significant at the 5% level and states that: " . . . it can be concluded at a 95% confidence that a statistical reduction in both the number of right-angle accidents and the number of total collisions did occur as a result of installing signal heads on mast arms and incorporating the one-second, all red interval." The author did not test whether the increase in left-turn accidents is statistically significant. (However, were he to use his inadequate method of testing, this too would be found statistically significant.)

> *Comment:* Three aspects of this claim deserve comment. First, in spite of this common misconception, a test of significance does not allow one to conclude that a change is real, or true, or that it occurred. Nor does it lead to the opposite conclusion. Unfortunately one cannot state briefly what is the conclusion of a statistical test. An attempt to provide a clear explanation is contained in Hauer (1996b).
>
> Second, a test of significance does not allow one to make attribution to cause. As has been asserted in this chapter, one could claim that "a reduction . . . did occur **as a result of** . . ." some treatment only if that treatment was the only change from 'before' to 'after'.
>
> Third, were one to follow the author's method and logic, one would have to conclude that the increase in left-turn accidents was also real and due to the same treatment. However, while the process by which a longer intergreen period reduces the number of right-angle accidents has been studied and is sufficiently well understood, it is not at all clear by what process a longer intergreen might induce an increase in left-turn accidents. To come to conclusions one must have evidence not only of the **product** (the estimated change in safety) but also of the **process** by which an treatment can affect safety. For left-turn accidents, the process is undocumented and the product too uncertain. Still, based on the data in hand, the most likely estimate of the effect of this treatment is a 32% increase in left-turn accidents.

What then can one learn from reviewing this article? First, that the study is of a Naive Before-After kind, and therefore one cannot decide what part of the estimated change in safety is due to the treatment (better signal head location and extension of the intergreen) and what part is due to the various unaccounted-for factors. Second, that some data were discarded on grounds that are either insufficient or insufficiently explained. Third, that regression-to-the-mean was likely at work and, if yes, a part of the large reduction in right-angle accidents would have been observed even if the intersections were left unmodified. Fourth, that tests of significance do not distinguish between 'true' and 'not true'. Fifth, that it is usually important to dwell on the process by which safety is influenced.

7.5 CHAPTER SUMMARY

By a Naive Before-After study one estimates how much safety has changed between the 'before' and the 'after' periods. However, one cannot say what part of this change is due to the treatment and what part is due to all other factors that have also changed during that time. Therefore, all written accounts of such studies should be accompanied by appropriate disclaimers. When such disclaimers are missing, the informed reader has to be able to provide them.

The statistical machinery used to assess how precisely the change in safety is estimated has been provided in this chapter. When the change in safety is large, the number of accidents needed is small enough to make a study practically feasible. However, changes in safety that amount to only a few percentage points may be practically undetectable; the number of accidents needed to detect such small changes is very large.

In the next two chapters we will seek ways to separate the effect of the treatment from the effect of other factors that affect safety and change with time. All attempts to do so come at a price. The price is the loss of statistical precision in estimation. Estimates may become more specific to the treatment the effect of which is sought, but will have a larger standard deviation than that obtained in a Naive Before-After study. It follows that when conclusions are to be drawn on the basis of accident counts, **the statistical precision of the Naive Before-After study is the limit of what is possible.** When coupled with the earlier observation that small changes in safety cannot be detected even in the study with the best precision, one must conclude that they are seldom detectible.

CHAPTER 8

IMPROVING PREDICTION I:

FACTORS MEASURED AND UNDERSTOOD

The problem with the Naive approach is that it cannot distinguish between what is the effect of the treatment, and what is the effect of the many other factors that have also changed from the 'before' to the 'after' period. The source of this problem is in the inadequacy of the 'prediction' component of the Naive method. To 'predict' is to guess what would have been the safety of the treated entities in the 'after' period had treatment not been applied. The essence of the Naive approach to prediction is the assumption that, was the treatment not implemented, the safety of an entity in the 'after' period would be what its safety in the 'before' period was. But this can only be true if traffic, precipitation, road user demography, vehicle fleet, economic conditions, safety measures, and a host of other factors were the same in both periods. Since this is seldom even approximately true, one has to seek ways to account for the safety effect of changes in these factors.

Many causal factors influence road safety. For our purposes they can be placed into one of two classes shown in Figure 8.1.

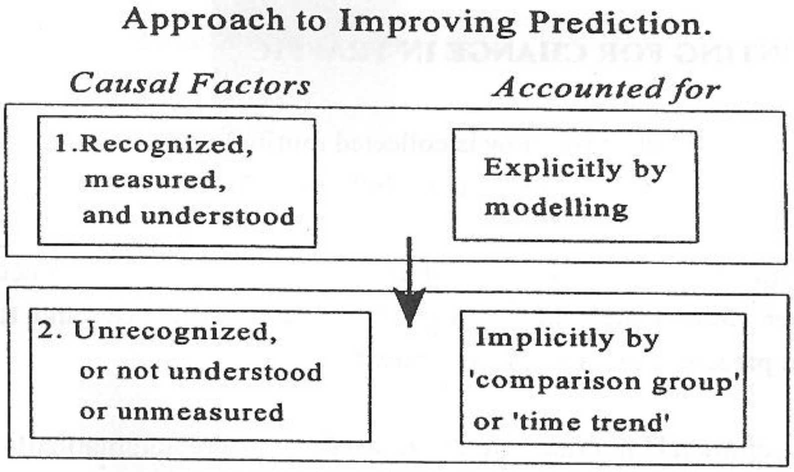


Figure 8.1. Sequence of corrections to account for the influence of change in causal factors.

To the first class belong causal factors the level of which can be estimated for the 'before' and 'after' periods, and the influence of which on safety can be modeled. It seems sensible to estimate the safety effect of factors in this class directly and explicitly. To the second class belong those causal factors that are presently unrecognized as affecting safety; factors suspected to affect safety but not measured; and other factors that, even if recognized and measured, can not be modeled because their influence on safety is at present ill-understood. To account for the influence of factors of this class it is necessary to resort to the use of a 'comparison group' or to the prediction of a trend from a time series of accident counts. These subjects will be discussed in the next chapter.

I think that it is in keeping with most scientific tradition to use explicit modeling to account for change in those factors that have been measured and the influence of which is sufficiently understood. Thus, e.g., in the laboratory one applies a 'temperature correction'; in precision distance taping one uses a 'tape sag correction' etc. Only after the prediction is corrected for the influence of the 'recognized-measured-understood' factors, should one account for the change in the remaining causal factors, those that are unrecognized, unmeasured and ill-understood. This sequence of actions is indicated by the arrow in Figure 8.1.

One important causal factor that changes between the 'before' and 'after' periods, and the influence of which on safety can be modeled, is 'traffic flow'. How to account for the influence of changes in traffic flow will be discussed in detail here. Other recognized-measured-understood factors can be dealt with similarly. Once this is done, it is possible to claim that whatever the estimated change in safety from the 'before' to the 'after' period, it is not due to change in factors that have been explicitly accounted for.

8.1 ACCOUNTING FOR CHANGE IN TRAFFIC

Information about traffic flow is collected routinely for a variety of purposes and therefore is often available for studies on road safety. In fact, if a Before-After study is planned, information about traffic flow in the 'before' and 'after' periods should always be secured. Since the effect of change in traffic flow on safety may be large, it is important to try to account for it directly and explicitly, even though the relationship between the traffic flow and the expected number of accidents is at present only imperfectly known.

It is a common belief that changes in traffic flow are automatically accounted for by using accident rates (such as accidents/vehicle kilometer) instead of accident frequencies (such as accidents/year). This belief is correct only if the expected accident frequency is directly proportional to traffic flow, which is apparently seldom the case. In so, what is commonly believed is usually inaccurate[1]. However, because the assumption of direct proportionality is so commonplace and may occasionally be a reasonable approximation, it will be the subject of Section 8.2. Dealing first with what is familiar also serves a didactic purpose. The general case when the proportionality assumption is not made will be discussed in Section 8.3.

Each entity has its specific 'before' and 'after' traffic flows. Therefore, the effect of change in traffic flow has to be accounted for separately for each entity. For this reason, the discussion below is about applying a correction for change in traffic flow to a single entity. In earlier chapters 'j' has been used to identify an entity. However, to keep the notation simple, r_d, K, L, π etc. will be used instead of the more complete notation $r_d(j)$, K(j) and L(j), $\pi(j)$.

Recall that Step 1 of the four-step (see Section 6.1) requires the prediction of what would have been the expected number of target accidents during the 'after' period had it been left without treatment, π. Figure 8.2 depicts the process of improving the prediction of a Naive study as a series of enhancements.

Both periods of same duration;
Traffic same in both periods; $\pi = \kappa$
Remaining factors same in both periods.

Traffic same in both periods;
Remaining factors same in both periods. $\pi = r_d \kappa$

Remaining factors same in both periods. $\pi = r_d r_{tf} \kappa$

Figure 8.2. Multiplicative corrections to the predictor π.

In Figure 8.2 assumptions are listed to the left and the corresponding predictor to the right. In each step of the series, one unsatisfactory assumption is removed and replaced by a **correction factor** in the predictor. Thus, were nothing to change from 'before' to 'after', and if both periods

1 More has been said on this in Section 3.3.

were of the same duration, there would be good reason to assert that $\pi=\kappa$; that is, that the number of accidents expected in the 'after' period (π) is the same as that in the 'before' period (κ). When the two periods are identical in all respects but different in duration, and if the ratio of their durations is r_d, the assertion should be modified to $\pi = r_d\kappa$. The question before us is: "how to modify the predictor of π if all conditions remained the same, except that the 'before' and 'after' durations are different and traffic flow has changed from 'B' in the 'before' period to 'A' in the 'after' period?". This will be done by using another multiplicative correction - the 'traffic flow correction factor'- denoted by r_{tf}. The subscript 'tf' stands for traffic flow. The r_{tf} is similar to 'duration correction' r_d. To illustrate the approach, consider the following numerical example[1].

Numerical Example 8.1. Resurfacing of one road section.

Consider a stretch of road on which during a 2-year 'before' period there were 30 wet-pavement accidents (w.p.a.'s) during 50 wet-pavement days. The road was then resurfaced. During the following 2-year 'after' period there were 40 w.p.a.'s in 40 wet-pavement days. The average daily traffic during a wet-pavement day of the 'before' period was 3928 vehicles; during the 'after' period, 4376 vehicles. The question is, how many wet-pavement accidents would be expected in the 'after' period had the road not been resurfaced.

Had nothing changed from 'before' to 'after', we would predict 30 w.p.a.'s for the 'after' period. Were the number of wet-pavement days the only difference, we would predict for the 'after' period $30\times(40/50)=24$ w.p.a.'s. But there was, in addition, the change in traffic. For the purpose of this illustrative example assume that wet-pavement-accidents are proportional to the amount of traffic. If so, we predict for the 'after' period $30\times(40/50)\times(4376/3928)=26.7$ w.p.a.'s.

In this example, the correction for traffic flow rests on the assumption that the expected number of accidents is proportional to traffic flow. The more general approach is to assume that the relationship between the expected number of target accidents and traffic flow (for entities similar to the treated one) is represented by a function f=f(flow). This will be the 'safety performance function'. A prototype safety performance function is shown in Figure 8.3. The

[1] In this and most other numerical examples the values were computed by a spreadsheet. Since rounded values of intermediate results are shown, minor discrepancies between the final result and the intermediate results used in its computation may be noted.

literature containing estimated safety performance functions for a variety of roads and intersection types is rapidly growing[1].

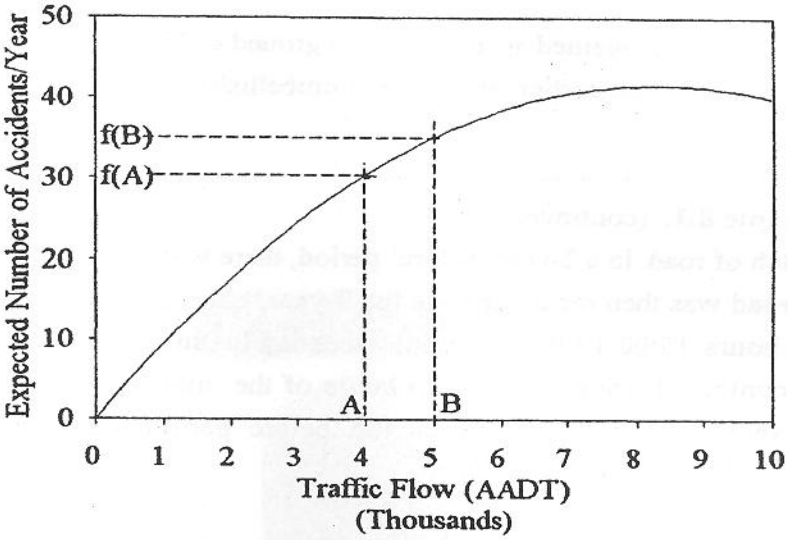


Figure 8.3. A safety performance function.

When traffic flow changes from flow=B to flow=A, the expected number of target accidents changes from f(B) to f(A). Define

$$r_{tf} \doteq f(A)/f(B) \qquad \ldots 8.1$$

The 'B' and 'A' are mnemonics for 'Before' and 'After'. Because the function represents the influence of traffic flow on accidents for entities that are similar to the treated one, we may assume that were the traffic flow on the treated entity to change from flow B to flow A, the expected number of accidents would change from κ (in the 'before' period) to $r_{tf}\kappa$ (in the 'after' period). Were also the duration of the two periods different, the predictor would be, in this case,

$$\pi = r_d \times r_{tf} \times \kappa \qquad \ldots 8.2$$

Equation 8.1 defines the multiplicative correction, in this case, for traffic flow. Equation 8.2 shows how it features in the predictor. In the next Section (8.2) I will show how this fits into the 'four-step'. The addition to the variance of estimation which is due to this new correction is examined in Section 8.3. How this variance is influenced by the short-duration traffic counts that are used to estimate traffic flow is explored in Section 8.4. Section 8.5 will pull all the strands together, and Section 8.6 will provide a summary.

---

[1] What is known about how accident frequency depends on traffic flow is discussed in Chapter 6 of Gartner, Messer and Rathi (1997?). For a review of many North American safety performance functions see Hauer and Persaud (1996).

8.2 THE TRAFFIC FLOW CORRECTION IN THE 'FOUR-STEP'

The approach is best explained against the background of Numerical Example 8.1 the bare bones of which were introduced earlier and which is embellished here.

Numerical Example 8.1. (continued).

On a stretch of road, in a 2-year 'before' period, there were 30 wet-pavement accidents (w.p.a.'s). The road was then resurfaced. In the 2-year 'after' period there were 40 w.p.a.'s. During the two hours 15:00-17:00 of a rainy weekday in June of the 'before' period, 572 vehicles were counted. During similar two hours of the 'after' period, 637 vehicles were counted. There were 50 wet-pavement days in the 'before' period and 40 in the 'after' period. The data are summarized in the table below.

Table 8.1. Data summary.

	Accident count	2-hour traffic	Wet-pavement days
Before	K=30	572	50
After	L=40	637	40

We wish to find the bottom-line estimates of δ, θ and their standard deviations. STEP 1 is to find estimates of λ and π. As in the Naive method, we will use L to estimate λ. The main task here is to predict: "What would have been the expected number of wet-pavement accidents in the 'after' period without resurfacing (π)." The phrase 'in the after period' means that what occurred in the 'before' period has to be modified to recognize that in the 'after' period there were fewer wet-pavement-days but more traffic. To account for the reduced number of wet-pavement days one may assume strict proportionality. This mimics the accounting for the different durations of the 'before' and 'after' periods already discussed in Section 7.1. Here the 'before' period is 50 days and the 'after' period 40 days. Thus, the duration correction r_d=40/50. Note that r_d is taken to be known without error. To account for the effect of the change in traffic flow, one has to know:

- **a.** how the number of wet-pavement accidents depends on the flow of traffic and
- **b.** what were the traffic flows B and A in the 'before' and 'after' periods.

For didactic purposes I will assume in this section that the expected number of accidents is proportional to traffic flow. In the next section I will show that, when the proportionality assumption holds,

$$\hat{r}_{tf} = \hat{A}_{avg} / \hat{B}_{avg} \quad \text{and} \qquad \ldots 8.3$$

$$VAR\{\hat{r}_{tf}\} = r_{tf}^2\,(v^2\{\hat{A}_{avg}\} + v^2\{\hat{B}_{avg}\}) \qquad \ldots 8.4$$

In this, $\hat{A}_{avg}$ and $\hat{B}_{avg}$ are the average traffic flows during the 'after' and 'before' periods and the coefficient of variation, v{.}, is a number that depends on the duration of the count and is gleaned from Table 8.7 given in Section 8.4.

The equations for Steps 1 and 2 for each entity are as in Table 8.2. The first row is the same as for the Naive method. The estimate $\hat{\pi}$ is based on Equation 8.2. In the 'derivations' at the end of this section I show how the expression for $V\hat{A}R\{\hat{\pi}\}$ is obtained. Steps 3 and 4 are, as always, given by Equations 6.1 to 6.4 in Table 8.3.

Table 8.2. Equations for STEPS 1 & 2 when safety is a linear function of traffic flow.

Estimates of Parameters	Estimates of Variances
$\hat{\lambda} = L$	$V\hat{A}R\{\hat{\lambda}\} = L$
$\hat{\pi} = r_d\,\hat{r}_{tf}\,K$	$V\hat{A}R\{\hat{\pi}\} = (r_d)^2\,[(\hat{r}_{tf})^2\,K + K^2\,V\hat{A}R\{\hat{r}_{tf}\}]$

Table 8.3. Equations for STEPS 3 & 4.

$$\delta \equiv \pi - \lambda \qquad \ldots 6.1$$

$$VAR\{\hat{\delta}\} = VAR\{\hat{\pi}\} + VAR\{\hat{\lambda}\} \qquad \ldots 6.2$$

$$\theta^* = (\lambda/\pi)/[1+VAR\{\hat{\pi}\}/\pi^2] \qquad \ldots 6.3$$

$$VAR\{\hat{\theta}\} \approx \theta^2[(VAR\{\hat{\lambda}\}/\lambda^2)+(VAR\{\hat{\pi}\}/\pi^2)]/[1+VAR\{\hat{\pi}\}/\pi^2]^2 \qquad \ldots 6.4$$

I now return to the numerical example.

Numerical Example 8.1. (continued).

STEP 1. $\hat{\lambda} = 40$ w.p.a.

$\hat{r}_{tf} = 637/572 = 1.114.$

$\hat{\pi} = (40/50)\times 1.114\times 30 = 26.73$ w.p.a.

Thus, we predict that were the road to remain without resurfacing, the number of wet-pavement accidents in the 'after' period (i.e., with the wet-pavement days and traffic flow of the 'after' period) would have been, on the average, 26.73.

STEP 2. So far we had r_d=40/50, $\hat{r}_{tf}$=1.114 and K=30. However, the magnitude of VAR{$\hat{r}_{tf}$} still remains to be estimated. How to do so will be discussed in Section 8.3. For the data of this example we will find that $\hat{VAR}\{\hat{r}_{tf}\}$=0.036. With this,

$\hat{VAR}\{\hat{\lambda}\} = 40$ w.p.a.^{2}

$\hat{VAR}\{\hat{\pi}\} = (40/50)^2[1.114^2\times 30+30^2\times 0.036]$

$=0.64[1.24\times 30+1.08\times 30] = 44.4$ w.p.a.^{2}.

The constituent parts of $\hat{VAR}\{\hat{\pi}\}$ in Table 8.2 can be now discerned and interpreted. If the number of wet-pavement days and traffic flow (as well as all other factors) is known to be the same in the 'before' and 'after' periods, then, $\hat{VAR}\{\hat{\pi}\}$ would have been K (30 in the numerical example). Was only the number of wet-pavement days to change (from 50 to 40), this would make the variance r_d^2 K (=0.64×30). Was traffic flow known to increase exactly by a factor r_{tf} (=1.114) the variance would be $r_d^2\ r_{tf}^2$ K (=0.64×1.24×30=23.8). However, as the change in traffic is not known exactly, Equation 8.4 must be used. This adds 0.64×1.08×30=20.7 to the variance. Thus, the fact that traffic has been estimated from a short duration count is seen to almost double the variance in this example. Having accomplished the tricky part, that of determining what $\hat{\pi}$ and $\hat{VAR}\{\hat{\pi}\}$ are, we can return to the comfort of routine to estimate δ, θ, and their standard deviations.

The equations in Table 8.2 pertain to one entity. Since in Numerical Example 8.1 only one stretch of road has been resurfaced, these equations could be applied directly. When there are several entities, Equations 6.5 and 6.6 apply.

Table 8.4. STEPS 1 & 2 with many entities.

$\lambda \doteq \Sigma\lambda(j)$	
$\pi \doteq \Sigma\pi(j)$	...6.5
$VAR\{\hat{\lambda}\} = \Sigma VAR\{\hat{\lambda}(j)\}$	
$VAR\{\hat{\pi}\} = \Sigma VAR\{\hat{\pi}(j)\}$	...6.6

The equations for $\hat{\lambda}$, $\hat{\pi}$, $V\hat{A}R\{\hat{\lambda}\}$, $V\hat{A}R\{\hat{\pi}\}$ are in Table 8.4. The estimates are in the bottom part of Table 8.5. These now serve in STEP 3 and 4.

Numerical Example 8.2. Resurfacing of several road sections.

During a construction season 3 road sections were resurfaced. Data for one of these were analyzed in Numerical Example 8.1 and the result of STEPs 1 and 2 are given in row 1 of Table 8.5. A similar data analysis (STEP 1 and STEP 2) for the other two road sections led to the results in rows 2 and 3.

Table 8.5. Data and sums.

j	$\hat{\lambda}(j)$ w.p.a*	$\hat{\pi}(j)$ w.p.a.	$V\hat{A}R\{\hat{\lambda}(j)\}$ w.p.a.^{2}	$V\hat{A}R\{\hat{\pi}(j)\}$ w.p.a.^{2}
1	40	26.73	40	44.5
2	82	92.15	82	62.6
3	16	21.17	16	20.9
Sums	138	140.05	138	128.00
	$\hat{\lambda}$	$\hat{\pi}$	$V\hat{A}R\{\hat{\lambda}\}$	$V\hat{A}R\{\hat{\pi}\}$

* wet-pavement accidents

STEP 3. Find estimates of δ and θ using Equations 6.1 and 6.3.

$\hat{\delta}$ = 140.05-138 =-2.05 wet-pavement accidents.

$\hat{\theta} = (138/140.05)/[1+128.00/140.05^2] = 0.979$

STEP 4. Find the estimates of $VAR\{\hat{\delta}\}$ and $VAR\{\hat{\theta}\}$ using Equations 6.2 and 6.4.

$\hat{\sigma}\{\hat{\delta}\} = \sqrt{(128.00+138)} = 16.3$ wet-pavement accidents.

$\hat{\sigma}\{\hat{\theta}\} = (0.979)(1/138+128.00/140.05^2)^{½}/[1+128.00/140.05^2] = 0.114.$

To recapitulate, the main aim of this section was to illustrate how the correction for traffic flow is accommodated in the four-step. Equations 8.3 and 8.4 give estimates of the traffic flow correction r_{tf} and its variance. How these two feature in the prediction $\hat{\pi}$ and its variance is shown in Table 8.2. The remaining steps remain unchanged. Where Equations 8.3 and 8.4 come from, and how to correct for traffic flow when proportionality is not a reasonable assumption is the subject of the next section.

Derivations.

Equations for STEPs 1 and 2 in Table 8.2.

The first line of Table 8.2 is as for the Naive method; λ is estimated by L. Since L is Poisson distributed, $VAR\{\hat{\lambda}\}=\lambda$ and this is again estimated by L, as in the Naive method. The parameter π is given in Equation 8.2 and is estimated by replacing r_{tf} and κ by their estimates $\hat{r}_{tf}$ and K. Note that the traffic flow correction r_{tf} is not known exactly because traffic has been counted only for two hours in each period. This is why the ^ notation is used.

To obtain the variance of $\hat{\pi}=r_d\,\hat{r}_{tf}\,K$, seeing that it is a function of the random statistically independent variables $\hat{r}_{tf}$ and K, the method of statistical differentials (Section 6.2) is used. $\partial\hat{\pi}/\partial\hat{r}_{tf}=r_d\,K$ and $\partial\hat{\pi}/\partial K=r_d\,\hat{r}_{tf}$. Evaluating these at their expected values and using Equation 6.7 leads to $VAR\{\hat{\pi}\}\approx r_d^2\,[r_{tf}^2\,VAR\{K\}+(E\{K\})^2\,VAR\{\hat{r}_{tf}\}]$ which is estimated by $r_d^2\,[\,\hat{r}_{tf}^2\,K+K^2\,\widehat{VAR}\{\hat{r}_{tf}\}]$.

8.3 THE ESTIMATION OF r_{tf}

To account for changes in traffic flow we need an estimate of the ratio r_{tf}. This ratio has been defined as f(A)/f(B) where A and B are the traffic flows during the 'after' and 'before' periods. The function 'f' links the expected number of target accidents and traffic flow and is called the 'safety performance function'. A safety performance function has been depicted in Figure 8.3.

Safety performance functions for various road and intersection types can be found in the literature (see, e.g., Gartner Messer and Rathi (to be published in 1997), Hauer and Persaud (1996)). A brief comment on how safety performance functions are obtained is in order. The researcher had data about target accidents and traffic flow on many entities of some kind[1]. After some exploration of the data by cross tabulations, graphical means, and mainly by trial and error, a certain functional form has been tentatively chosen because it seems to fit the data best. In such a function, the expected accident frequency depends on traffic flow and on some unknown constants called 'parameters'. Thus, e.g., if κ denotes the expected accident frequency, one might conclude at the end of the exploratory data analysis that the polynomial function κ=f(traffic flow)=α(traffic

[1] More often than not, the researcher has data about several additional characteristics (such as lane width and grade for roads, or traffic control and sight distance for intersections). These too are part of the function. In the special case when the terms for the various characteristics are multiplicative, and remain unchanged, their value is immaterial since they cancel out of the ratio $\hat{r}_{tf}$. More will be said on this in Part III.

flow)+β(traffic flow)^2 is a sensible choice. In this, α and β are the unknown parameters. Next the parameters are estimated so as to best fit the observed data. After examining how the data fit the function, one may revisit the choice of functional form.

Once a safety performance function has been chosen and its parameters have been estimated, one then assumes[1] that it represent a cause-effect relationship. That is, when on such entities traffic flow[2] changes from B to A, the number of target accidents is expected to change from f(B) to f(A). Therefore, when estimates $\hat{A}$ and $\hat{B}$ of the traffic flows A and B are available, the r_{tf} can be estimated. In accord with the Definition 8.1, the estimate $\hat{r}_{tf}$ is given by

$$\hat{r}_{tf}=f(\hat{A}_{avg})/f(\hat{B}_{avg}) \qquad \ldots 8.5$$

In this, 'f' is the safety performance function and $\hat{A}_{avg}$ and $\hat{B}_{avg}$ are estimates of the average flows during the 'after' and 'before' periods respectively. These estimates are based on traffic counts that are usually taken on a few days of a year. In the special case when the proportionality assumption holds, that is, when the safety performance function is of the form $\alpha\times$(traffic flow), Equation 8.5 turns into Equation 8.3 that has been used in the preceding section.

For STEP 2 we also need to have an estimate of VAR$\{\hat{r}_{tf}\}$ (see, for example, the equations in Table 8.2). Note that the difference between A_{avg} and its estimate $\hat{A}_{avg}$ causes a difference between $f(A_{avg})$ and $f(\hat{A}_{avg})$ in the numerator of Equation 8.5. The difference between B_{avg} and its estimate $\hat{B}_{avg}$ causes a similar difference in the denominator. Jointly these two make for a difference between r_{tf} and $\hat{r}_{tf}$. This is measured by VAR$\{\hat{r}_{tf}\}$. In the previous section I have given an expression for the linear case (Equation 8.4) and promised to explain later how it has been obtained. Detailed derivations are provided at the end of this section. It now can be shown that, generally,

$$\text{VAR}\{\hat{r}_{tf}\} \approx r_{tf}^2\,[c_A^2\,\text{VAR}\{\hat{A}_{avg}\}\,/\,f^2(A_{avg}) + c_B^2\,\text{VAR}\{\hat{B}_{avg}\}\,/\,f^2(B_{avg})] \qquad \ldots 8.6$$

In this, c_A and c_B denote the derivatives of 'f' with respect to traffic flow at A_{avg} and B_{avg}. The application of Equation 8.6 to two common forms of 'f' is given below.

---

[1] This is a problematic assumption. The genesis of the safety performance function are data about many entities that carry different flows. The assumption implies that the same relationship holds on each entity as the flow on it changes.

[2] The traffic flow may be 'hourly flow', 'daily flow', 'annual average daily flow' and the like. These flows vary during the 'before' and 'after' periods. Therefore, in principle, the ordinates f(A) and f(B) which define r_{tf} are 'average ordinates'. Some, as yet unpublished, analysis indicates that the difference between the average ordinate or the ordinate for the average flow is likely to be small.

If the safety performance function is assumed to be linear, that is, if f(flow)=α×(flow), Equation 8.6 turns into

$$VAR\{\hat{r}_{tf}\} \approx r_{tf}^2\,(v^2\{\hat{A}_{avg}\}+v^2\{\hat{B}_{avg}\}) \qquad \ldots 8.7$$

where v{.} is the ratio of the standard deviation of a random variable and its mean (the 'coefficient of variation'). (Equation 8.7 has been introduced earlier as equation 8.4 without showing its origin). How estimates of v{.} can be obtained is discussed in Section 8.4.

If the safety performance function is of the form f(flow)=α×(flow)$^{\beta}$, then

$$VAR\{\hat{r}_{tf}\} \approx r_{tf}^2\beta^2(v^2\{\hat{A}_{avg}\}+v^2\{\hat{B}_{avg}\}) \qquad \ldots 8.8$$

We are now in a position to summarize in Table 8.6.

Table 8.6. Summary of estimates for r_{tf} and VAR{$\hat{r}_{tf}$}.

Safety Performance Function	Estimate of r_{tf}	Estimate of VAR{$\hat{r}_{tf}$}
f(flow)	$\hat{r}_{tf}=f(\hat{A}_{avg})/f(\hat{B}_{avg})$	$VAR\{\hat{r}_{tf}\}\approx r_{tf}^2[c_A^2VAR\{\hat{A}_{avg}\}/f^2(A_{avg})+c_B^2VAR\{\hat{B}_{avg}\}/f^2(B_{avg})]$
α×(flow)	$\hat{r}_{tf}=\hat{A}_{avg}/\hat{B}_{avg}$	$VAR\{\hat{r}_{tf}\}\approx r_{tf}^2(v^2\{\hat{A}_{avg}\}+v^2\{\hat{B}_{avg}\})$
α×(flow)$^{\beta}$	$\hat{r}_{tf}=(\hat{A}_{avg}/\hat{B}_{avg})^{\beta}$	$VAR\{\hat{r}_{tf}\}\approx r_{tf}^2\beta^2(v^2\{\hat{A}_{avg}\}+v^2\{\hat{B}_{avg}\})$

To use these equations we have to face the empirical question of what are the coefficients of variation of the estimates of the average flows A_{avg} and B_{avg} when these are based on traffic counts. Answering this question is the next task.

Derivations.

Equation 8.6.

We wish to find VAR{$\hat{r}_{tf}$}. Let f(A_{avg}) and f(B_{avg}) be the ordinates for the average flows of the 'after' and the 'before' periods respectively. We have defined the ration $r_{tf} \doteq f(A_{avg})/f(B_{avg})$. Estimates $\hat{A}_{avg}$ and $\hat{B}_{avg}$ of these average flows are obtained from traffic counts. We estimate r_{tf} by

$$\hat{r}_{tf}=f(\hat{A}_{avg})/f(\hat{B}_{avg}) \qquad \ldots 8.5$$

The random variable $\hat{r}_{tf}$ is a function of the random variables $\hat{A}_{avg}$ and $\hat{B}_{avg}$. To find its variance requires repeated application of the method of statistical differentials discussed in Section 6.2. First consider the random variable $\hat{r}_{tf}$ to be a function of the independent random variables $f(\hat{A}_{avg})$ and $f(\hat{B}_{avg})$. Using Equation 6.7 on the function defined by Equation 8.5

$$VAR\{\hat{r}_{tf}\} \approx r_{tf}^2\ [VAR\{f(\hat{A}_{avg})\}/f^2(A_{avg}) + VAR\{f(\hat{B}_{avg})\}/f^2(B_{avg})]$$

To understand this equation it may help to examine the term $VAR\{f(\hat{A}_{avg})\}/f^2(A_{avg})$. Recall that $f(A_{avg})$ is the expected number of accidents when traffic flow is A_{avg}. Thus, $VAR\{f(\hat{A}_{avg})\}$ is the variance of the expected number of accidents. It is caused by the fact that the estimated flow $\hat{A}_{avg}$ differs from the true A_{avg} because it is based on a short traffic count. To flesh the equation out, I have to suggest how to estimate $VAR\{f(\hat{A}_{avg})\}$ and $VAR\{f(\hat{B}_{avg})\}$. Using the method of Equation 6.7 again, let c_A and c_B denote the derivatives of 'f' with respect to traffic flow and evaluated at A_{avg} and B_{avg}. Then,

$$VAR\{f(\hat{A}_{avg})\} = c_A^2\ VAR\{\hat{A}_{avg}\}$$
$$VAR\{f(\hat{B}_{avg})\} = c_B^2\ VAR\{\hat{B}_{avg}\}$$

With this, the equation turns into the general form listed as Equation 8.6 earlier:

$$VAR\{\hat{r}_{tf}\} \approx r_{tf}^2\ [c_A^2\ VAR\{\hat{A}_{avg}\}/f^2(A_{avg}) + c_B^2\ VAR\{\hat{B}_{avg}\}/f^2(B_{avg})] \qquad \ldots 8.6$$

Equation 8.7.

At this point the form of the function 'f' has to be declared. The most common assumption is that the safety performance function is linear, that is, that $f(\text{flow})=\alpha\times\text{flow}$. If so, the derivatives $c_A=c_B=\alpha$. In this case,

$$\begin{aligned} VAR\{f(\hat{A}_{avg})\}/f^2(A_{avg}) &= c_A^2\ VAR\{\hat{A}_{avg}\}/f^2(A_{avg}) \\ &= \alpha^2\ VAR\{\hat{A}_{avg}\}/\alpha^2(A_{avg})^2 \\ &= VAR\{\hat{A}_{avg}\}/(A_{avg})^2 \\ &\doteq v^2\{\hat{A}_{avg}\}\ . \end{aligned}$$

The ratio of the standard deviation of a random variable and its mean is called its 'coefficient of variation' and is usually denoted by v{.}. Similarly,

$$\begin{aligned} VAR\{f(\hat{B}_{avg})\}/f^2(B_{avg}) &= VAR\{\hat{B}_{avg}\}/(B_{avg})^2 \\ &= v^2\{\hat{B}_{avg}\} \end{aligned}$$

Therefore, when f(flow)=$\alpha\times$(flow), Equation 8.6 turns into

$$VAR\{\hat{r}_{tf}\} \approx r_{tf}^2\,(v^2\{\hat{A}_{avg}\} + v^2\{\hat{B}_{avg}\}) \qquad ...8.7$$

Equation 8.8.

Another commonly used form of the function 'f' is f(flow)=α(flow)$^{\beta}$. The derivative of 'f' with respect to flow is then f(flow)$\times\beta$/flow. In this case,

$$\begin{aligned} VAR\{f(\hat{A}_{avg})\}/f^2(A_{avg}) &= c_A^2\, VAR\{\hat{A}_{avg}\}/f^2(A_{avg}) \\ &= \beta^2\, VAR\{\hat{A}_{avg}\}/(A_{avg})^2 \\ &= \beta^2\, v^2\{\hat{A}_{avg}\} \end{aligned}$$

and similarly,

$$VAR\{f(\hat{B}_{avg})\}/f^2(B_{avg}) = \beta^2\, v^2\{\hat{B}_{avg}\}$$

Now Equation 8.6 takes on the form

$$VAR\{\hat{r}_{tf}\} \approx r_{tf}^2\, \beta^2\, (v^2\{\hat{A}_{avg}\}+v^2\{\hat{B}_{avg}\}) \qquad ...8.8$$

8.4 COEFFICIENTS OF VARIATION FOR AADT ESTIMATES

Usual practice is to count traffic for a road section or intersection during a relatively short period and to 'factor'[1] this count up to give an estimate of the Annual Average Daily Traffic - the AADT. How this sampling and 'factoring' is done varies somewhat among jurisdictions. There seem to be two main ways for doing so. One is to identify one of the several permanent counting stations available as the best match for the time-profile of the short-duration traffic count. The short-duration count is then multiplied by the ratio of the annual flow at the permanent counting station and its flow during the same time as that of the short-duration traffic count. This is divided by 365 to estimate the AADT. In an (as yet unpublished) study Vallurupalli found that the percent coefficient of variation can be estimated by:

$$v=1+7.7/(\text{number of count-days})+1650/AADT^{0.82} \qquad ...8.9$$

Thus, if on a road section with an AADT of about 5000 traffic was counter for three days, the coefficient of variation can be expected to be $1+7.7/3+1650/5000^{0.82}$=5.1%.

---

[1] A 'short period' may be less than a day or up to a dozen or so days. The 'factoring up' entails making from this sample an average for a whole year.

The other way of 'factoring' short-duration traffic counts to AADT is to obtain a series of multipliers that account for time-of-day, day-of-week and month-of-year during which the traffic was counted. These multipliers are specific to a few road types. The multipliers are developed from data obtained at so called 'Permanent Counting Stations', where, as the name attests, traffic is counted all the time. My aim is to illustrate how these estimation practices translate into the coefficients of variation[1] (v). The procedure used for this purpose is of the second type and has been described by Phillips (1979). It distinguishes between four types of sites:

1. Urban and/or commuter sites;
2. Non-recreational low flow (rural) sites;
3. Rural long distance sites and
4. Recreational sites.

To illustrate, if the count was on a 'rural long distance' site, took place on a weekday in June, and was 16 hours long, it has to be multiplied by 391 (See Phillips, 1979, Table 7). These so called 'M-factors' are used to calculate the annual flow. If only a two-hour count between 15:00 and 17:00 took place, it has to be first multiplied by 6.41 to make it into a 16-hour count (see Phillips, 1979, Table 6). The obtainable accuracy of the estimated annual flow is described by the coefficient of variation the percent-values of which are in Table 8.7. The use of the coefficients of variation is illustrated in Numerical Example 8.3.

Table 8.7. Percent coefficients of variation.

Month	Length of Count	Urban/ Commuter	Non-recreational Low flow	Rural Long-dist.	Recreational
MAY	2 hours	11	14	13	19
	4 hours	9	12	11	17
	6 hours	8	11	10	16
	16 hours	6	10	9	16
JUNE	2 hours	12	15	12	17
	4 hours	9	13	10	16
	6 hours	8	12	9	15
	16 hours	7	11	8	14
SEPT.	2 hours	13	14	13	22
	4 hours	10	12	11	20
	6 hours	9	11	10	20
	16 hours	8	8	10	9

[1] Recall that the coefficient of variation is the ratio of the standard deviation and the mean.

Numerical Example 8.3. Coefficients of variation.

On a rural, long-distance site, 572 vehicles were counted during 15:00-17:00 on a weekday in June. From this, the estimated annual flow is 572×6.41×391=1433609 vehicles per year. (The M-factor is 391 and the multiplier for counting 2 hours and not 16 is 6.41. These were mentioned earlier. For more detail, see Phillips, 1979). Dividing by 365 we obtain 3928 vehicles on an average day. For this case, from Table 8.7, the coefficient of variation is 0.12 (or 12%). Therefore the standard deviation of the estimated annual flow is 1433609×0.12=172033 and the standard deviation of the estimated average daily flow is 3928×0.12=471.

In the next section I pull together all the strands of analysis and illustrate by numerical examples how to account for the effect of traffic flow.

8.5 ILLUSTRATIONS AND DISCUSSION

To account for the change in traffic flow from the 'before' to the 'after' period we need estimates of r_{tf} and of VAR$\{\hat{r}_{tf}\}$ (see Table 8.2). I have explored separately the case when the expected number of accidents is taken to be proportional to traffic flow and the case when the relationship is taken to be non-linear. How to proceed when proportionality is a reasonable assumption is illustrated in Numerical Example 8.4.

Numerical Example 8.4. Accidents are proportional to AADT.

I continue here the story line of Numerical Example 8.3 where 572 vehicles have been counted during two-hours of the 'before' period, and 637 during a two-hour count in the 'after period'. The coefficients of variation were estimated to be 0.12. In this example it is assumed that the expected number of accidents is proportional to the annual average daily traffic (AADT). By the equation in Table 8.6 (or, Equations 8.3 and 8.4), $\hat{r}_{tf}$=637/572=1.114 and $\widehat{\text{VAR}}\{\hat{r}_{tf}\} \approx 1.114^2\times[2\times0.12^2]$ =0.036. These were the values used earlier in Numerical Example 8.1.

Most research finds that the relationship between the expected number of accidents and traffic flow is non-linear. How to proceed in this case is illustrated in Numerical Example 8.5.

Numerical Example 8.5. Accidents are proportional to $AADT^{\beta}$.

The most commonly used functional relationship is $f(flow)=\alpha(flow)^{\beta}$. Assume that for the target accidents and kinds of road of interest it has been established that $\beta=0.8$. Using the data from the previous example and the equations in Table 8.6, $\hat{r}_{tf}=(637/572)^{0.8}=1.114^{0.8}=1.090$ and $\hat{VAR}\{\hat{r}_{tf}\}= 1.114^2\times0.8^2[2\times0.12^2]=0.022$.

Note that it matters what the functional form of the safety performance function is. The results of the two numerical examples are compared in Table 8.8.

Table 8.8. Comparison of results.

Example	f(flow)	$\hat{r}_{tf}$	$\hat{VAR}\{\hat{r}_{tf}\}$
8.4	$\alpha\times$(flow)	1.114	0.036
8.5	$\alpha\times(flow)^{\beta}$	1.090	0.022

Since the modelling of the relationship between traffic and accidents is evolving, and in the future many functional forms may be encountered, it may be useful to show how to deal with any specified functional form of f(flow).

Numerical Example 8.6. Accidents are a specified function of flow.

Assume now that on facilities of this kind, the relationship between the expected number of accidents per unit of time and traffic flow is given by, say, $f(flow)=flow/100-0.006(flow/100)^2$. For this safety performance function use has to be made of the general Equations 8.5 and 8.6 as I show below. On the basis of the two-hour 'before' count of 572 vehicles $AADT_{before}$ has been estimated (in Numerical Example 8.3) to be $572\times6.41\times391/365=3928$. Similarly, using the two-hour 'after' count of 637 vehicles, $AADT_{after}$ is estimated to be $637\times6.41\times391/365=4374$. Thus, $f(3928)=3928/100-0.006(3928/100)^2=30.02$ accidents/unit of time and $f(4374)=32.26$ accidents/unit of time. Therefore, by the equation in the first row of Table 8.6 (or Equation 8.5), $\hat{r}_{tf}=32.26/30.02=1.075$.

For Equation 8.6 we have to estimate the derivatives c_A and c_B. The derivative of this safety performance function with respect to flow is $1/100-(0.012/100^2)\times flow$. Using flow=4374, $c_A=0.00475$; using flow=3928 and $c_B=0.00529$. Since the coefficient of variation for these two-hour counts is 0.12, the standard deviations of the two estimates of AADT are $0.12\times3928=471$ and $0.12\times4374=525$. Using these values in Equation 8.6, (also listed in the first row of Table 8.6) $\hat{VAR}\{\hat{r}_{tf}\}\approx(1.075)^2\,[(0.00475\times525/32.26)^2 + (0.00529\times471/30.02)^2] = 0.015$. In sum, $\hat{r}_{tf} = 1.075$ and $\hat{VAR}\{\hat{r}_{tf}\}\approx0.015$.

I now return to Numerical Example 8.1 that was first introduced in Section 8.1. There the assumption was that the expected number of accidents is proportional to traffic flow. I now examine the same case when f(flow) is non-linear.

Numerical Example 8.1. (continued from Section 8.1).

Consider again the stretch of road on which in a 2-year 'before' period there were 30 w.p.a.(wet-pavement accidents). The road was then resurfaced. In the 2-year 'after' period there were 40 w.p.a.. During the two hours 15:00-17:00 of a rainy weekday in June of the 'before' period, 572 vehicles were counted. During a similar two hours of the 'after' period, 637 vehicles were counted. There were 50 wet-pavement days in the 'before' period and 40 in the 'after' period. The information added here is that the literature indicates that on roads like this the expected number of accidents is proportional to $(AADT)^{0.8}$. We wish to estimate δ, θ and the standard deviations of these estimates. As noted in Numerical Example 8.5 and in Table 8.8, $\hat{r}_{tf}=1.114^{0.8}=1.090$ and $\hat{VAR}\{\hat{r}_{tf}\}=0.022$. With this we can go through the four-step.

STEP 1. Find estimates of λ and π.

	$\hat{\lambda}=40$	(Table 8.2)
	$\hat{r}_{tf}=1.090$	(Table 8.6 or Equation 8.5)
	$\hat{\pi}=(40/50)\times1.090\times30=26.16$ w.p.a.	(Table 8.2 or Equation 8.2)

STEP 2. Find estimates of VAR$\{\hat{\lambda}\}$ and VAR$\{\hat{\pi}\}$.

	$\hat{VAR}\{\hat{\lambda}\}=40$ (w.p.a.)^2	(Table 8.2)
	$\hat{VAR}\{\hat{\pi}\}=(4/5)^2[1.090^2\times30+30^2\times0.022]$	(Table 8.2)
	$=35.4$ (w.p.a.)^2	

STEP 3. Find estimates of δ and θ.

$$\hat{\delta}=26.16-40=-13.84 \text{ w.p.a.}$$
$$\hat{\theta}=(40/26.16)/[1+35.4/26.16^2]=1.45.$$

STEP 4. Find the estimates of VAR$\{\hat{\delta}\}$ and VAR$\{\hat{\theta}\}$.

$$\hat{\sigma}\{\hat{\delta}\}=\sqrt{(35.4+40)}=8.7 \text{ w.p.a.}$$
$$\hat{\sigma}\{\hat{\theta}\}=(1.45)\sqrt{(1/40+35.4/26.16^2)}/[1+35.4/26.16^2]=0.38$$

When in Section 8.1 accidents were assumed proportional to flow, the corresponding estimates were $\hat{\delta}=-13.27$ wet-pavement accidents, $\hat{\theta}=1.41$, $\hat{\sigma}\{\hat{\delta}\}=9.2$ wet-pavement accidents and $\hat{\sigma}\{\hat{\theta}\}=0.39$.

8.6 CHAPTER SUMMARY

The task of Chapter 8 was to account for the effect on safety of factors that are measured in the 'before' and 'after' periods and the influence of which is understood. In truth, only one such factor has been considered - traffic flow. However, if one could specify 'performance functions' for other factors, such as g(road user demography), h(unemployment), I(vehicle fleet), each giving the expected number of target accidents as a function of these factors, the approach outlined in this chapter would still apply.

The phrase 'to account for a factor' means to predict what would have been the change in safety from the 'before' to the 'after' period due to a measured change of some factor such as traffic flow. The ability to so account or predict, rests on two pieces of knowledge. The first is the knowledge of how the expected number of target accidents depends on the factor of interest; in our case, on traffic flow. Thus, one needs to have a 'safety performance function' linking the expected accident frequency to traffic flow. The amount of traffic flow is only imperfectly known. In our case, estimates of traffic flow are obtained from short-duration traffic counts. Accordingly, the second piece of knowledge needed is an assessment of how accurately the factor of interest is known.

My analysis of $\widehat{VAR}\{\hat{r}_{tf}\}$ is deficient. The assumption is that the safety performance function is known, and that we are only unsure of r_{tf} because A_{avg} and B_{avg} are estimated from short-duration traffic counts. In fact we are not sure what the correct functional form of the safety performance function is, nor are we sure about the parameters of the function. These two additional sources of uncertainty should also be accounted for. Thus, in reality, $\widehat{VAR}\{\hat{r}_{tf}\}$ is larger than what our analysis indicates it to be.

I started this chapter by making a distinction between those causal factors that are recognized- measured-understood and the other group of casual factors that are unrecognized-and/or-unmeasured-and/or-ill understood. How to deal with the first group has been discussed in this chapter. The next chapter is devoted to the discussion of how to account for change in the causal factors of the second group.

CHAPTER 9

IMPROVING PREDICTION II: USING A COMPARISON GROUP

To maintain the sense of direction and progress it may help to review what has been accomplished so far, and to say what is to be attempted next. To estimate the safety effect of a treatment, it is necessary to predict what would have been the safety of an entity (or of a group of entities) in the 'after' period had treatment not been applied. There are many ways to predict, as has been discussed in Chapter 5, and there are correspondingly many ways to do a Before-After study. One way to predict is to assume that the number of target accidents in the 'after' period is expected to be the same as that in the 'before' period. This is the essence of the Naive Before-After study. This assumption is obviously at variance with reality; many factors that influence safety change as time passes. As a result, a Naive study cannot distinguish between what is caused by the treatment and what is caused by the many other influences (listed at the beginning of Chapter 7). One can predict better than by making such a Naive assumption.

To predict better it is necessary to account for the influence of a variety of causal factors that change with time. These were said to belong to two classes. In the first class are factors that are known to affect safety, the magnitude of which is measured both 'before' and 'after' treatment, and the effect of which on safety is understood sufficiently well. How to account for their influence has been discussed in Chapter 8. To the second class of causal factors belong those we do not recognize as affecting safety, those recognized but unmeasured, and those whose influence on safety is not well understood. These causal factors are the subject of the present chapter. I will examine one method of prediction that accounts for the influence of the unrecognized, and/or unmeasured, and/or ill-understood factors. The method makes use of a 'comparison group'.

I believe that one will predict best by first accounting for the causal factors of the first class and doing so by explicit modeling. Only after that has been done, should one account for the influence of all the remaining causal factors. However, this is not current practice. At present it is customary to account for the influence of all factors by making use of the comparison group device. Sections 9.1 to 9.5 are written to suit this common practice. The modifications that are necessary if the measured and understood factors are accounted for before a comparison group is used, are examined in Section 9.6.

The central idea of using a comparison group is simple: Identify a group of entities that remained untreated, and that are similar to the treated entities (or entity). The treated entities form the 'treatment group'. The untreated entities are the comparison group. Both are represented in Figure 9.1. The hope is that the change from 'before' to 'after' in the safety of the comparison group is indicative of how safety on the treatment group would have changed. This hope is based on two assumptions.

Assumption a. That the sundry factors that affect safety have changed from the 'before' to the 'after' period in the same manner on both the treatment and the comparison group, and

Assumption b. That this change in the sundry factors influences the safety of the treatment and the comparison group in the same way.

To be precise, let 'before' and 'after' periods be those of some treated entity. Define:

$r_C \doteq$ to be the '**comparison ratio**'; the ratio of expected number of 'after' to the expected number of 'before' target accidents on the comparison group.

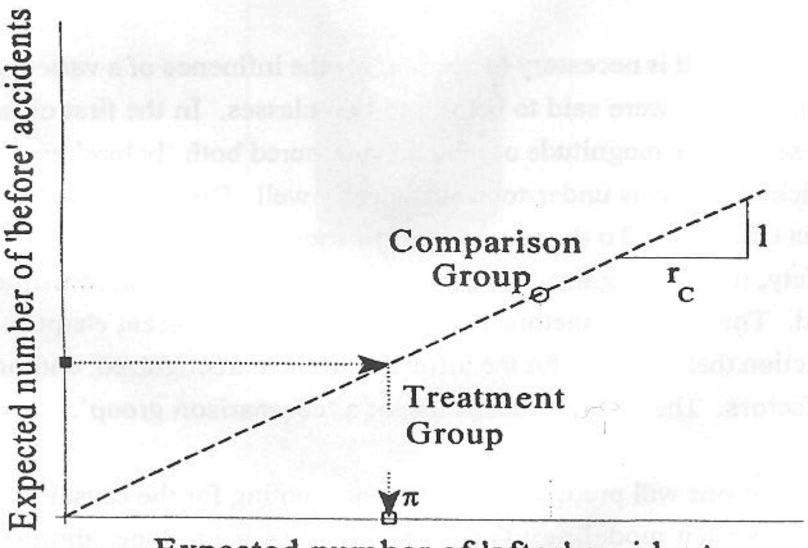


Figure 9.1. The definition of the 'comparison ratio' and its use for prediction.

The expected accident numbers for the comparison group are the ordinate and the abscissa of the circle in Figure 9.1. The slope of the line joining the circle to the origin is r_C. It is hoped that, had the treatment group been left untreated, the ratio of expected number of 'after' to 'before' target accidents on it would also be r_C. If the hope is true, then one may **predict:**

$$\pi = (\text{expected number of 'before' target accidents on the treatment group}) \times r_C. \qquad ...9.1^1$$

The expected accident numbers for the treatment group are shown in Figure 9.1 as squares. The empty square is the prediction π. This hope, as embodied in Equation 9.1, is the essence of the Comparison Group Method. I will call it the 'C-G method'. When, in a study, use is made of the 'C-G method', it will be called a 'C-G study'.

The idea of the C-G method is deceptively similar to that of randomized experiments which are popular in agriculture, medicine, the social sciences and in many other fields of research. However, there is a crucial distinction. In a randomized experiment, the decision as to which entities get treated and which are left as 'control' is made **at random.** Therefore, were the experiment repeated a very large number of times, each time with a random assignment of entities to treatment and 'control', then, in the limit, the influence of the sundry causal factors on both groups of entities would tend to be equal. As a consequence, when the assignment to treatment is 'at random', it is legitimate to speak of a '**statistical experiment**' which involves a '**control group**'. In contrast, when the assignment of entities to the treatment group is not made at random, then, even if both groups of entities are very large, they will differ systematically with respect to some causal factors. Therefore, even with large groups of entities there is no assurance that the expected number of accidents in the treatment group (had treatment not been administered) would have changed in the same manner as in the comparison group. For this reason, when entities are **not assigned to treatment at random,** one may not speak of an 'experiment' or of a 'control group'. To mark the distinction, it is prudent to always speak of '**observational studies**' (not of experiments) and of '**comparison groups**' (not of control groups). The main ideas of using a comparison group in an observational study and some of the problems associated with it, are now introduced by a numerical example[2].

---

[1] If a treated entity j has its own comparison group, $\pi(j)$=(expected number of 'before' target accidents on entity j)$\times r_C(j)$.

[2] For further discussion see the endnote.

Numerical Example 9.1. Comparison Group (C-G) Method for R.I.D.E..

In Numerical Example 7.1 it has been said that: "In one of five police districts in Metropolitan Toronto an enforcement program to reduce impaired driving (R.I.D.E.) has been introduced. One year before the program started there were 173 alcohol-related injuries. During the first year that the R.I.D.E. program was in place, there were 144 alcohol-related injuries." Within the framework of a Naive Before-After study, the estimates of δ and θ obtained in Numerical Example 7.1 were said to reflect the influence of a variety of factors, not only that of R.I.D.E..

Consider now one of the remaining four police districts where R.I.D.E has not been implemented to be a comparison group. Call it district X. The count of alcohol-related injuries in district X was: 225 'before' and 195 'after' (Vingilis et al., 1979). The task is to predict what would have been the expected number of alcohol-related injuries in the treated district, had it been left without treatment.

The workings of the sundry factors in comparison district X are estimated to have reduced the expected number of target accidents in the ratio 195/225=0.867. Thus $\hat{r}_C$=0.867. Were identical sundry influences at work in the treated district, and if they influenced its expected alcohol-related injuries in the same manner, we would predict the same reduction; that is, from 173 to 173×0.867=149.9. Thus $\hat{\pi}$=149.9 alcohol-related injuries.

A few problems now need to be recognized. First, the assumption behind this prediction method is that the sundry factors (except for R.I.D.E.) exerted identical influences in police district X and in the treatment district. This is unlikely to be exactly true because somewhat different causal factors and at different levels may have acted on two different police districts; there may have been a major snowstorm in one, and a big rock concert in the other. In addition, it is unlikely that the effect of increased enforcement under R.I.D.E., and the attendant publicity, were neatly confined to the boundaries of precisely one police district - the treatment group - and had no influence in another police district in the same metropolitan area.

The second complication is, that I could have chosen any one of the four untreated police districts W, X, Y, Z or any combination thereof as a comparison group. Each district had a different estimated comparison ratio; each choice would have led to a different estimate of π as is shown in Table 9.1. The question is, which choice would have been best.

Table 9.1. Possible comparison groups.

Police District	W	X	Y	Z	(W+X+Y+Z)
'Before' [a.r.i*]	289	225	179	204	897
'After' [a.r.i.]	259	195	193	223	870
Comparison ratio	0.896	0.867	1.078	1.093	0.970
$\hat{\pi}=173\times\hat{r}_c$ [a.r.i.]	155.0	149.9	186.5	189.1	167.8

* alcohol-related injuries

Third, the information we have are the accident counts in the comparison group, while the comparison ratio is defined in terms of expected values. It follows that the comparison ratio is subject to the usual imprecision of estimation. It is therefore necessary to determine how accurately can π (and thus δ and θ) be ascertained by the C-G method. An answer to this question is needed for the correct interpretation of the results and also for the planning of a C-G study.

These problems will be addressed later in this chapter. The next task is to show how the C-G method fits into the general four-step schema.

9.1 Statistical analysis

In an observational Before-After study involving a treatment and a comparison group, let K, L, M and N in Table 9.2 denote the accident counts that correspond to the row and column headings. The expected values of these accident counts are denoted by the corresponding Greek letters κ, λ, μ and ν.

Table 9.2. Accident counts and expected values.

	Treatment Group	Comparison Group
Before	K, κ	M, μ
After	L, λ	N, ν

Details of the standard four steps[1] are given at the end of this section under 'derivations'. Here I only dwell on part of the first step because of the important assumption which it embodies and a new variable that needs to be introduced, as a result.

---

[1] Step 1. Find estimates of λ and π; Step 2. Find estimates of VAR$\{\hat{\lambda}\}$ and VAR$\{\hat{\pi}\}$; Step 3. Find estimates of δ and θ; Step 4. Find estimates of VAR$\{\hat{\delta}\}$ and VAR$\{\hat{\theta}\}$.

The C-G method is based on the hope that, in the absence of treatment, the ratio of the expected number of target accidents 'before' and 'after' would be the same in the treatment and the comparison groups. To bring the hope into the orbit of explicit analysis I have to distinguish between the ratio r_C defined for the comparison group and a parallel but distinct ratio for the treatment group. Thus, define:

$r_C \doteq \nu/\mu$ to be the ratio of the expected accident counts for the comparison group, and

$r_T \doteq \pi/\kappa$ to be the corresponding ratio for the treatment group[1] ... 9.2

The aforementioned hope can now be expressed as an equation. The hope is that

$$r_T = r_C \quad \text{or, equivalently, that} \quad r_C/r_T = 1 \qquad \text{... 9.3}$$

From the definition of r_T it follows that

$$\pi = r_T\,\kappa \qquad \text{... 9.4}$$

Therefore, if the assumption in Equation 9.3 is true, then it is also true that

$$\pi = r_C\,\kappa \qquad \text{... 9.5}$$

Since r_C in Equation 9.5 can be estimated from the number of accidents in the comparison group (M and N in Table 9.2), and κ can be estimated by the number of accidents in the treatment group in the 'before' period (K), π can be estimated.

It is now apparent, that the hope expressed by Equation 9.3 is the foundation of the C-G method (see note at the end of this section). Just as the Naive method rests on an assumption[2], so does the C-G method. Just as the assumption underlying the Naive method is never exactly true, so is the basic assumption of the C-G method never quite correct. One cannot argue with conviction that, because of some external similarity between the treatment and the comparison groups, the assumption is 'almost true' or 'likely to be true'. Plentiful data show this argument to be invalid (see Hauer et al., 1991 and Hauer, 1991). A more coherent argument needs to be constructed.

The only defensible argument that I can muster to justify the use of a comparison group in an observational study is empirical or inductive. Namely, if one can show that in a time series of past values r_T and r_C were sufficiently similar then, cognizant of the usual limitations of all inductive arguments, one may hope that the past similarity also held for that specific value of r_C that is used

---

[1] Note that r_T is defined as π/κ, not as λ/κ, since treatment was involved. The numerator π captures 'what would have been had the treatment not been implemented'.

[2] The assumption is that nothing has changed from 'before' to 'after'.

in a specific 'C-G study'. However, if this is the argument on which the C-G method rests, one must allow in the analysis for the possibility that the assumption $r_C/r_T=1$ is not exactly true in any specific 'C-G study'. It is therefore necessary to consider the ratio r_C/r_T to be a random variable which on different occasions takes on different values. Consistent with the usual statistical terminology associated with 2×2 tables such as Table 9.2 the ratio r_C/r_T will be called the 'odds ratio' and given the symbol ω (see, e.g., Fleiss, 1981).

$$\omega \doteq r_C/r_T \qquad \ldots 9.6$$

Imagine now a long time series of accident counts for a treatment and a comparison group of entities. From these accident counts form a time series of 2×2 tables such as Table 9.2. The four accident counts K, L, M and N in each table are for the same set of entities, except that at no time has the treatment been applied to the entities of the treatment group. In each such table a specific value of ω holds. Thus, for each group of treatment and comparison entities there is a **time series of ω's**. Any such sequence of ω's has a mean $E\{\omega\}$ and a variance $VAR\{\omega\}$. For a comparison group to be considered legitimate, it must fulfill the requirement that $E\{\omega\}=1$. If this requirement is not met, the past r_C must have been systematically larger or systematically lesser than the corresponding r_T. This would completely negate the basic premise of the 'C-G study'.

Now the road to the statistical analysis of a C-G study is clear of conceptual obstacles. The results for Steps 1 and 2 are listed in Table 9.3. Derivations are provided at the end of this section. Table 9.4 shows the usual equations for STEPs 3 and 4.

Table 9.3. Estimates for STEPs 1 and 2 in a 'C-G study'.

Estimates of Parameters STEP 1	Estimates of Variances STEP 2
$\hat{\lambda}=L$	$V\hat{A}R\{\hat{\lambda}\}=L$
$\hat{r}_T=\hat{r}_C=(N/M)/(1+1/M)\approx N/M$	$V\hat{A}R\{\hat{r}_T\}/r_T^2\approx 1/M+1/N+V\hat{A}R\{\omega\}$
$\hat{\pi}=\hat{r}_T K$	$V\hat{A}R\{\hat{\pi}\}\approx\hat{\pi}^2[1/K+V\hat{A}R\{\hat{r}_T\}/r_T^2]$

Table 9.4. STEPs 3 and 4.

$$\delta \equiv \pi-\lambda \qquad \ldots 6.1$$
$$VAR\{\hat{\delta}\} = VAR\{\hat{\pi}\} + VAR\{\hat{\lambda}\} \qquad \ldots 6.2$$
$$\theta^* = (\lambda/\pi)/[1+VAR\{\hat{\pi}\}/\pi^2] \qquad \ldots 6.3$$
$$VAR\{\hat{\theta}\} \approx \theta^2[(VAR\{\hat{\lambda}\}/\lambda^2)+(VAR\{\hat{\pi}\}/\pi^2)]/[1+VAR\{\hat{\pi}\}/\pi^2]^2 \qquad \ldots 6.4$$

In the Naive method $\hat{\pi}=K$. In the C-G method, $\hat{\pi}=\hat{r}_T K$. The purpose of $\hat{r}_T$ (or of its replacement $\hat{r}_C$) is to account for the effect of change in various uncontrolled causal factors. The consequence of the addition of $\hat{r}_T$ is that the statistical precision with which π can be estimated must now depend not only on VAR{K} but also on VAR{$\hat{r}_T$}, as is evident in the last row of Table 9.3. Two conclusions follow. First, that the statistical precision in the estimation of π that is attainable by the C-G method is less than that achievable by the Naive method, because an additional source of variance has been introduced. The C-G method yields estimates that are specifically of the safety effect of the treatment, unlike the Naive method which does not distinguish between the effect of the treatment and of all other causal factors that change in time. However, this improvement is bought at the expense of an increase in the variance of the bottom line estimates. The second conclusion pertains to guidance on how to choose among candidate comparison groups. If several comparison groups that all have an average ω near 1 are available, one should choose that group, which makes the end result most precise. Inspection of Table 9.3 indicates that one should **select that comparison group for which 1/N+1/M+V$\hat{A}$R{ω} is smallest**. To illustrate the use of these results, two linked numerical examples are used. In the first I show how to choose the best comparison group, in the next example I use the chosen comparison group to estimate the effect of R.I.D.E..

Numerical Example 9.2. Choice of comparison group.

For the statistical analysis of this and the next example we use the data of Numerical Example 9.1 and of Table 9.1. Table 9.1 contains accident counts from four police districts considered as possible comparison groups for the district in which R.I.D.E. was implemented. The task here is to decide which of the available comparison groups to use. Only then can one embark on the four-step procedure. Suppose that from a history of alcohol-related injury accident counts for each of the five police districts, all have an average ω of about 1. Estimates of VAR{ω} are in Table 9.5. (The procedure for obtaining such estimates will be discussed in Section 9.3.

Table 9.5. Estimates of VAR{ω}.

	Police District				
	W	X	Y	Z	W+X+Y+Z
V$\hat{A}$R{ω}	0.0032	0.0028	0.0046	0.0041	0.0055

Thus, e.g., for comparison group W, using the accident counts from Table 9.1, V$\hat{A}$R{$\hat{r}_T$}/r_T^2= 1/M+1/N+V$\hat{A}$R{ω}= 1/289+1/259+0.0032=0.0105. For comparison groups X, Y, Z and (W+X+Y+Z) similar computations lead to values 0.0124, 0.0154, 0.0135 and 0.0078. It follows that the composite comparison group W+X+Y+Z will give the best estimate of π. Note that it is best because it has the largest accident counts, not because it has the smallest VAR{ω}. Having decided on the comparison group, (that is, M=897 and N=870), the usual four-step can commence.

Numerical Example 9.3. R.I.D.E. with comparison groups.

STEP 1. Find $\hat{\lambda}$ and $\hat{\pi}$.

Table 9.6 gives the data about alcohol related injury counts (a.r.i.). By the equation for $\hat{\pi}$ in Table 9.3, without R.I.D.E. we would expect $173\times[(870/897)/(1+1/897)]=173\times0.969 =167.6$ a.r.i.

Table 9.6. Alcohol-related injury counts.

	Treatment	Comparison
Before	K=173	M=897
After	L=144	N=870

Thus:

$$\hat{\lambda}=144 \text{ a.r.i}; \quad \hat{r}_T=\hat{r}_C=0.969; \quad \hat{\pi}=173\times0.969=167.6 \text{ a.r.i}.$$

Note that the biased estimator $\hat{r}_C$ is 870/897=0.970 while the approximately unbiased estimator is 0.969. In this case, the difference hardly matters.

STEP 2. Find $\hat{VAR}\{\hat{\lambda}\}$ and $\hat{VAR}\{\hat{\pi}\}$.

As in earlier examples, $VAR\{\hat{\lambda}\}$ is estimated by L to be 144 a.r.i.^{2}. From Table 9.3 $\hat{VAR}\{\hat{r}_T\}/r_T^2=1/897+1/870+0.0055=0.0078$ and therefore, $\hat{VAR}\{\hat{\pi}\}=167.6^2[1/173+0.0078] =380.5$ [a.r.i.]^2. Thus,

$$\hat{VAR}\{\hat{\lambda}\}=144 \text{ (a.r.i.)}^2 \text{ and } \hat{VAR}\{\hat{\pi}\}= 380.5 \text{ (a.r.i.)}^2.$$

STEP 3. Find $\hat{\delta}$ and $\hat{\theta}$.

Making use of results in Steps 1 and 2, $\hat{VAR}\{\hat{\pi}\}/\hat{\pi}^2=380.5/167.6^2=0.014$. Using Equations 6.1 and 6.3 in Table 9.4,

$$\hat{\delta}=167.6-144=23.6 \text{ a.r.i. and } \hat{\theta}=(144/167.6)/1.014=0.848$$

STEP 4. Find the estimates of $VAR\{\hat{\delta}\}$ and $VAR\{\hat{\theta}\}$.

Using Equations 6.2 and 6.4 in Table 9.4,

$$\hat{\sigma}\{\hat{\delta}\}=\sqrt{(380+144)}=22.9 \text{ a.r.i and } \hat{\sigma}\{\hat{\theta}\}=0.848\times(1/144+0.014)^{1/2}/1.014=0.120$$

Thus, after accounting for the sundry influences which have changed from the 'before' to the 'after' period, we estimate that R.I.D.E. brought about a reduction of 23.6 alcohol-related injuries in one year. The standard deviation of this estimate is 22.9. This amounts to a 15.2% reduction with a standard deviation of 12%.

It is interesting to compare these results with what has been obtained earlier by a Naive study (Table 7.4) and the results of the attempt to take the time-trend into account (Section 7.2). Table 9.7 is a tangible manifestation of the fact that different methods give different results because different assumptions are made about how to predict 'what would have been'.

Table 9.7. Comparison of results obtained by different methods.

	Comparison Group	Naive	Time trend
$\hat{\delta}$	23.6 a.r.i.	29 a.r.i.	
$\hat{\theta}$	0.85	0.83	0.88
$\hat{\sigma}\{\hat{\delta}\}$	22.9 a.r.i.	17.8 a.r.i.	
$\hat{\sigma}\{\hat{\theta}\}$	0.120	0.09	

Note that the standard deviation of the estimates obtained by the Naive method are smaller than those obtained by the comparison-group (C-G) method. At first glance this may appear strange since the C-G method was intended to be an improvement on the Naive method. The improvement is there, but it is in terms of separating the effect of R.I.D.E. from the effect of the many other factors that also changed between the 'before' and the 'after' periods. This improvement is obtained at the expense of the statistical precision which is measured by the standard deviation. In this sense the results in Table 9.7 are typical.

Computations of this kind are best done in a simple spreadsheet program. A possible layout is shown in Table 9.8. Once the accident counts are entered into cells B3, B4, C3, C4 and the estimate of the variance of the odds ratio into cell B5, all the other numbers are automatically calculated. The column entitled 'Formula' list the formulae in cells C8 to C16.

Table 9.8. A prototype spreadsheet.

	A	B	C	Formula
1		INPUT:		
2		Treatment	Comparison	
3	Before	173	897	
4	After	144	870	
5	$\hat{VAR}\{\omega\}$	0.0055		
6				
7		OUTPUT:		
8	Step 1	$\hat{\lambda}=$	144	B4
9		$\hat{r}_C=$	0.969	(C4/C3)/(1+1/C3)
10		$\hat{\pi}=$	167.61	B3*C9
11	Step 2	$\hat{VAR}\{\hat{\lambda}\}=$	144	C8
12		$\hat{VAR}\{\hat{\pi}\}=$	380.49	C10^2*(1/B3+1/C3+1/C4+B5)
13	Step 3	$\hat{\delta}=$	23.6	C10-C8
14		$\hat{\theta}=$	0.848	(C8/C10)/(1+C12/C10^2)
15	Step 4	$\hat{\sigma}\{\hat{\delta}\}=$	22.9	@SQRT(C11+C12)
16		$\hat{\sigma}\{\hat{\theta}\}=$	0.120	C14*@SQRT(C11/C8^2+C12/C10^2

Derivations.

The derivations below pertain to the equations in Table 9.3

1. $\hat{\lambda}$ and $\hat{VAR}\{\hat{\lambda}\}$.

The first row of Table 9.3 is as for the Naive method. The random variable L is taken to be Poisson distributed with mean and variance λ. Therefore L estimates λ and also the variance of its estimator.

2. $\hat{r}_C$.

By Equation 9.2 I have defined the comparison ratio $r_C = \nu/\mu$. Thus, r_C could be estimated by the ratio N/M. However, this would be a biased estimator. The reason is the same as that already discussed in connection with Equation 6.3 (in Section 6.1). The derivation of an approximately unbiased estimator has been the subject of the illustration in Section 6.2. Applying the same argument here would lead to $E\{N/M\} \approx [\nu/\mu][1+VAR\{M\}/\mu^2]$. For the Poisson distribution $VAR\{M\}=\mu$ and $VAR\{M\}/\mu^2=1/\mu$. Therefore, to have an approximately unbiased estimate of r_C, N/M has to be divided by (1+1/M). When M is large, this correction is small and can be omitted.

3. $\widehat{VAR}\{\hat{r}_T\}/r_T^2$.

I called r_C/r_T the 'odds ratio' and designated it by the letter ω. There can be no assurance that for a specific treatment, comparison group, and a certain pair of 'before' and 'after' periods, $\omega=1$. It is prudent to think that over a sequence of many different 'before' and 'after' periods, different values of ω would materialize. Thus, over a long sequence of 'before' and 'after' periods ω is thought to have a mean $E\{\omega\}$ and a variance $VAR\{\omega\}$. A comparison group would be perfect if for every pair of 'before' and 'after' periods ω was 1 and thus $VAR\{\omega\}=0$. However, any real comparison group will have $VAR\{\omega\}>0$. How to estimate the $VAR\{\omega\}$ from a time series of accident counts will be discussed in Section 9.3. Here we wish to establish how $VAR\{\hat{r}_T\}/r_T^2$ depends on the random variation in the accidents counts M and N and the variance of ω.

From the definition of the odds ratio it follows that $r_T=r_C/\omega=(\nu/\mu)/\omega$. We estimate r_T by $\hat{r}_T \approx (N/M)/\hat{\omega}$. To this I apply method of statistical differentials (Section 6.2, Equation 6.7). Thus, $\partial\hat{r}_T/\partial N=1/(\hat{\omega}M)=\hat{r}_T/N$. When evaluated at the expected values and squared, it is r_T^2/ν^2. When multiplied by the variance of N this leads to $r_T^2VAR\{N\}/\nu^2$. Similarly the other two terms in Equation 6.7 will be $r_T^2VAR\{M\}/\mu^2$ and $r_T^2VAR\{\omega\}/E^2\{\omega\}$. Since for the Poisson distribution $VAR\{M\}=\mu$ and $VAR\{N\}=\nu$, $VAR\{\widehat{r^T}\} \approx r_T^2/\mu+r_T^2/\nu+ r_T^2VAR\{\omega\}/E\{\omega\}^2$ or

$$VAR\{\hat{r}_T\}/r_T^2 \approx 1/\mu+1/\nu+VAR\{\omega\}/E\{\omega\}^2.$$

For a comparison group to be considered legitimate we must have $E\{\omega\}=1$. Therefore the division by $E\{\omega\}$ can be omitted. To obtain the expression in row two of Table 9.3, replace μ by M, ν by N, and $VAR\{\omega\}$ by its estimate $\widehat{VAR}\{\omega\}$ (see Section 9.3).

4. $\hat{\pi}$.

By Equation 9.4, $\pi=r_T\,\kappa$. Since K estimates κ, $\hat{\pi}=\hat{r}_TK$.

5. $\widehat{VAR}\{\hat{\pi}\}$.

Here the method of statistical differentials (Section 6.2, Equation 6.7) is applied to $\hat{\pi}=\hat{r}_TK$. As in (3) above, the partial derivatives of $\hat{\pi}$ when evaluated at the expected values are π/r_T and π/κ.

Squared and multiplied by the corresponding variances we get $\pi^2 VAR\{\hat{r}_T\}/r_T^2$ and $\pi^2 VAR\{K\}/\kappa^2 = \pi^2/\kappa$. From here,

$$VAR\{\hat{\pi}\} \approx \pi^2[1/\kappa + VAR\{\hat{r}_T\}/r_T^2]$$

This gives guidance on which of several candidate comparison groups to choose. Only $VAR\{\hat{r}_T\}/r_T^2$ depends the choice of the comparison group. Since the objective is to minimize the $VAR\{\hat{\pi}\}$, that comparison group which yields the smallest $VAR\{\hat{r}_T\}/r_T^2$ should be chosen. (See Numerical Example 9.2). To obtain the expression in Table 9.3, replace π by $\hat{\pi}$, κ by K and $VAR\{\hat{r}_T\}/r_T^2$ by the expression in the second row of Table 9.3.

9.2 STUDY DESIGN CONSIDERATIONS FOR THE 'C-G method'

As noted earlier, 'study design' means one thing when one is to embark upon a randomized experiment and another when one plans an observational study. Preparations for an observational study are more like scavenging. The questions are: "what data can be found?" and "must I look for more?" In the design of a Naive study (see Section 7.3) the decisions were about the number of entities for the treatment group, and the duration of the 'before' and 'after' periods. In designing a C-G study, an additional decision is required, namely, which of the available candidate comparison groups should be used.

Thus, the principal choices to be made in the design of a 'C-G study' are five:

1. Select the **size of the treatment group** in terms of the number of target accidents expected to occur on its entities during a typical year of the 'before' period.

2. Select the **duration of the 'before' period.** The product of the choices in 1 and 2 is an indication what κ to aim for.

3. Decide what is the appropriate **duration of the 'after' period.** If the effect of the treatment can be assumed not to change with time, different considerations apply than when the effect of the treatment must be assumed to change with time. The choices made in 2 and 3 determine r_d (the ratio of the durations of the 'after' to 'before' periods). The ratios r_C and r_T are largely determined by the r_d. At the time of study design r_C and r_d may be taken to be approximately equal.

4. Postulate what is the anticipated **index of effectiveness** (θ) of the treatment that is to be estimated by the study. Once this is done, λ has also been determined (since $\lambda=\theta\pi$, $\pi=r_T\kappa$ and $r_T \approx r_d$, $\lambda \approx \theta r_d \kappa$).

5. Select the **comparison group.** This determines the expected number of accidents occurring during the 'before' period (μ) and the variance of the odds ratio (VAR{ω}). Having decided on r_d earlier, $\nu=r_d\mu$.

To make these 'study design decisions', we need to know how they will affect the precision of the bottom-line estimates. For this purpose, VAR{$\hat{\theta}$} needs to be expressed as a function of the relevant 'decision variables'. In the derivations at the end of this section I show that

$$\mathrm{VAR}\{\hat{\theta}\} \approx (\theta/r_d+\theta^2)/\kappa+\theta^2[(1/r_d+1)/\mu+\mathrm{VAR}\{\omega\}] \qquad \ldots 9.7$$

This equation links the VAR{$\hat{\theta}$} that we are aiming for, to the 'decision variables'. Note that the first summand in Equation 9.7 is the same as for the Naive method (compare to Equation 7.3). Therefore the graphs shown in Chapter 7 can be used here too, as will be demonstrated in a numerical example. The second summand is the addition to the variance of $\hat{\theta}$ that is due solely to the use of a comparison group. Once again it becomes clear that the use of a comparison group is tied to a precision cost. While helping to account for the effect of factors that have changed from the 'before' to the 'after' period, the use of a comparison group causes VAR{$\hat{\theta}$} for a C-G study to be larger than that of a Naive study. This increase depends in part on the number of accidents occurring on the entities of the comparison group, and in part on how variable the odds ratio is. How Equation 9.7 is used for the statistical design of a C-G study is illustrated in numerical example 9.4.

Numerical Example 9.4. Size of comparison group.

As in Numerical Example 7.3, consider a treatment that is thought to reduce the expected number of accidents by 10% (i.e., θ=0.9). We have found there, that to get $\sigma\{\hat{\theta}\} \approx 0.05$ in a Naive study with one year long 'before' and 'after' periods, some 700 'before' accidents were needed. Now, in order to eliminate from the estimate the effect of change in factors other than the treatment, it has been suggested to make use of a comparison group.

Numerical Example 9.4. Size of comparison group. (continued)

Question 1: What is the smallest comparison group to be considered to get $\sigma\{\hat{\theta}\}\approx 0.05$?

Answer 1: The requirement that $\sigma\{\hat{\theta}\}\leq 0.05$ translates into $VAR\{\hat{\theta}\}\leq 0.0025$. Because the 'before' and 'after' periods are of equal duration, $r_d=1$. If κ was very large, the first summand in 9.7 would be small. Furthermore, if the comparison group were an ideal one (i.e., if $VAR\{\omega\}=0$) we would need to ensure only that $2\theta^2/\mu \leq 0.0025$. Under such ideal circumstances, $\mu \geq 2\times 0.81/0.0025 = 648$. Thus, to attain the desirable accuracy of estimation under the best possible circumstances, the smallest comparison group needed is still a very large one.

A comparison group has been identified which (in a one year 'before' period) has M=2000 accidents and $V\hat{A}R\{\omega\}=0.003$. (How to obtain estimates of $VAR\{\omega\}$ will be discussed next, in Section 9.3).

Question 2: What must be the number of 'before' accidents on the treatment group to ensure that $\sigma\{\hat{\theta}\}\leq 0.05$?

Answer 2: Assuming that $\omega\approx 1$, the comparison group will contribute $0.9^2(2/2000+0.003) = 0.0032$ to $VAR\{\hat{\theta}\}$. But, we are trying to make sure that $VAR\{\hat{\theta}\}\leq(0.05)^2= 0.0025$. Thus, the aim can not be attained unless one can find a better comparison group.

Question 3: Suppose that a different comparison group was found with M=5000 and $V\hat{A}R\{\omega\}=0.001$. Can a C-G study attain the desired accuracy? If yes, what must be the expected number of 'before' accidents, κ?

Answer 3: The new comparison group contributes $0.9^2\times(2/5000+0.001)=0.0011$ to the $VAR\{\hat{\theta}\}$. This leaves only $0.0025-0.0011=0.0014$ to be the contribution of the treatment group to the variance.. From Figure 7.4a or Equation 7.3, or the first summand in Equation 9.7, the required number of 'before' accidents is now $(0.9/1+0.9^2)/0.0014\approx 1200$.

Equation 9.7 is a succinct representation of the link between the decisions that need to be made and the anticipated precision of the result of a future 'C-G study'. It can therefore be used to plan a 'C-G study' as has been illustrated. I find that a spreadsheet such as in Table 9.9 is often a more practical tool for study design than equation 9.7. Entering various possible numbers in the 'INPUT' part will show what $\sigma\{\hat{\theta}\}$ will obtain. A few such tries will tell the user what affects precision and how; what precision is obtainable and what plans are practical. The formulae that go into the computation of the eight 'OUTPUT' numbers are also shown.

Table 9.9. Spreadsheet for designing a 'C-G study'.

	A B C D E	F			
1	Designing a Comparison Group study.				
2					
3	**INPUT**				
4	'Before' accidents/year on treated system=	1200			
5	Number of 'before' years=	1			
6	Number of 'after' years=	1			
7	'Before' accidents/year on comparison system=	5000			
8	variance of odds ratio VAR{ω}=	0.001			
9	Expected % reduction $(1-\theta)\times 100$=	10			
10					
11					
12	**OUTPUT**				
13		treatment	compariso n	**Formulae**	
14	Before	1200	5000	+F4*F5	+F7*F5
15	After	1080	5000	+F4*F6*(1-F9/100)	+F7*F6
16					
17					
18	'After' accidents without treatment π =	1200		D14*E15/E14	
19	Estimate of VAR{$\hat{\pi}$}=	3216		+F18^2*(1+D14*F8+D14/E14+D14/E15)/d14	
20	Estimated index of effectiveness θ=	0.9		+D15/F18	
21	Estimated $\sigma(\hat{\theta})$=	0.051		(F20^2*(1/D15+F19/F18^2))^0.5	

What distinguishes a Naive study from a C-G study is the comparison group. Early in this chapter I showed that the choice of the comparison group often determines what conclusions are reached. This has been underscored dramatically by Welbourne (1989) who reexamined the Swedish evaluation of the safety effect of the Daytime Running Light (DRL) legislation mentioned earlier in Section 4.2 (Andersson et al., 1981). The assumption in their evaluation was that neither nighttime nor single-vehicle accidents can be affected by DRL. Thus, one could have used as the M and N (see Table 9.2) the count of single-vehicle daytime accidents (sd), or single-vehicle nighttime accidents (sn), or multivehicle nighttime accidents (mn) or a combination thereof. The estimates of DRL effectiveness shown in Table 9.10 would have been obtained. The estimate of the last row has been published.

Table 9.10. Alternative estimates of DRL effect.

Comparison Group Chosen	Percent Reduction in Target Accidents
sd	+6.0
mn	-2.1
sn	-7.6
sd+mn	+1.4
mn+sd	-4.0
sn+sd	+0.4
sd+mn+sn	-0.7
sd(mn/sn)	+10.8

Because the choice of the comparison group is often crucial, choosing it well is important. To choose well one has to have clear guidance available. I have said earlier that for a comparison group to be legitimate, the mean of a time series of ω's must be close to 1. Of those candidate comparison groups that meet this dictum, it is best to use that comparison group for which the sum $1/N+1/M+\hat{VAR}\{\omega\}$ is the smallest. This is perhaps too definitive and captures only the quantitative and inductive element of what guidance may be given. A more qualitative consideration has to do with the purpose for which the comparison group is used, namely, to account for the influence of change in various unaccounted-for factors. Therefore, for a comparison group to be legitimate, there should be reasons to believe that the change in these factors from 'before' to 'after' were the same in both groups of entities (stated as 'Assumption a' earlier) and that their influence was similar in the treatment and the comparison groups (stated earlier as 'Assumption b'). What then should guide the choice of the comparison group? In practical terms, a comparison group should meet the following requirements:

1. The 'before' and 'after' periods for the treatment and the comparison group should be the same; if not, 'Assumption a' is violated.

2. There should be reason to believe that the change in the factors influencing safety, whatever these may be, is similar in the treatment and comparison groups; this too is necessary for 'Assumption a' to hold.

3. The accident counts should be sufficiently large and;

4. When a sequence of sample odds ratios is calculated from historical accident counts, their sample mean is close to 1 and their variance is small.

Most guidance comes from requirements 1 and 3 and 4. Requirement 2 is little more than a declaration of intent. The factors to be accounted for are not recognized, or unmeasured or their effect is not understood. Therefore, one can judge similarity only vaguely[1].

In this and the preceding section I have used estimates of VAR{ω} without giving a procedure for its estimation. This will be the subject of the next section.

Derivation.

Equation 9.7:

By Equation 6.4, $VAR\{\hat{\theta}\} \approx \theta^2[(VAR\{\hat{\lambda}\}/\lambda^2)+(VAR\{\hat{\pi}\}/\pi^2)]/[1+VAR\{\hat{\pi}\}/\pi^2]^2$. For study design purposes the bias correction $[1+VAR\{\hat{\pi}\}/\pi^2]^2$ can be safely neglected. Therefore, I will use only

$$VAR\{\hat{\theta}\} \approx \theta^2[(VAR\{\hat{\lambda}\}/\lambda^2)+(VAR\{\hat{\pi}\}/\pi^2)]$$

For the Poisson random variable $\hat{\lambda}=L$ (see the first row of Table 9.3). Thus, $VAR\{\hat{\lambda}\}/\lambda^2=1/\lambda$. From the derivations at the end of the previous section we have: $VAR\{\hat{\pi}\} \approx \pi^2[1/\kappa+VAR\{\hat{r}_T\}/r_T^2]$ and $VAR\{\hat{r}_T\}/r_T^2 \approx 1/\mu+1/\nu+VAR\{\omega\}/E\{\omega\}^2$. Joining the two expressions, and taking $E\{\omega\} \approx 1$,

$$VAR\{\hat{\pi}\}/\pi^2 \approx [1/\kappa+1/\mu+1/\nu+VAR\{\omega\}].$$

Therefore,

$$VAR\{\hat{\theta}\} \approx \theta^2[1/\lambda+1/\kappa+1/\nu+1/\mu+VAR\{\omega\}]$$

The decisions variables are κ, r_d, θ and μ. Using $\lambda=\theta\pi$, $\pi=r_T\kappa$ and $r_T \approx r_d$ at the study design stage, we can write $\lambda=\theta r_d\kappa$. Replacing λ by $\theta r_d\kappa$ and ν by $r_d\mu$ we get:

$$VAR\{\hat{\theta}\} \approx \theta^2[1/\theta r_d\kappa+1/\kappa+1/r_d\mu+1/\mu+VAR\{\omega\}]=(\theta/r_d+\theta^2)/\kappa+\theta^2[(1/r_d+1)/\mu+VAR\{\omega\}].$$

---

[1] In Section 9.4 I have examined the effect on safety of replacing STOP signs by YIELD signs. In each of three cities such replacements were made at several intersections over a period of 6 years. While the duration of the 'before' and 'after' periods may have been (roughly) the same, the calendar years covered by the 'before' and 'after' periods differed from site to site. Therefore, different comparison ratios should have been used. Also, because not all sites were in the same city, it would have been difficult to argue that the change in relevant factors was common to all three cities. Therefore, to use a single comparison group for the three cities, as I have done, is likely to violate requirement 'b'.

9.3 ESTIMATION OF VAR{ω}

In Section 9.1 I have defined the odds ratio ω as r_C/r_T. I have argued there, that it is in the nature of treatment and comparison groups that for any particular 'before-after' period-pair, ω does not exactly equal 1, even if the comparison group is an ideal one. Only in a long sequence of before-after period-pairs, in each of which a different value of ω holds, will the mean E{ω} be equal 1. Since the assumption that $r_T=r_C$ (that is, that ω=1) is used in forming the prediction π, VAR{$\hat{\pi}$} must be influenced by the VAR{ω}. Inasmuch as the precision of the bottom-line estimates depends on the precision with which π is estimated, we need to estimate VAR{ω}. The aim of the present section is to suggest a procedure for the estimation of VAR{ω}. To introduce the subject, consider the following setting:

Numerical Example 9.5: A comparison group for New Brunswick.

Fatal accidents in New Brunswick are the target of some treatment. In a search for a comparison group, fatal accidents in Saskatchewan are considered. The historical record shows that fatal accidents in these two provinces track each other fairly closely. This is shown in Figure 9.2. The question is what are the mean and the variance of the odds ratio ω.

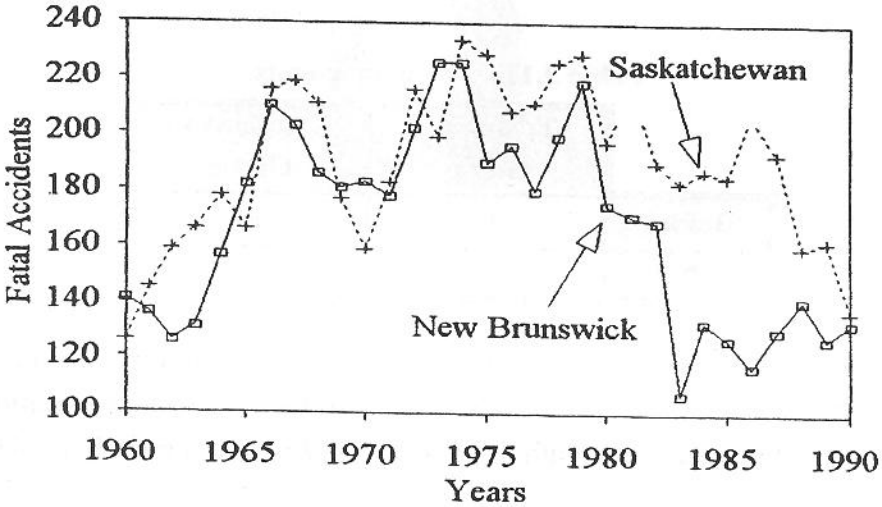


Figure 9.2. Time series of fatal accidents.

The task is to use data such as that shown in Figure 9.2 in order to examine how the odds ratio ω varied in this period. Proofs and detailed argumentation are provided in the 'derivations' at the end of this section; only the main results are stated here. Consider a time series of accident counts for a treatment group and a candidate comparison group similar to that in Figure 9.2. A 'time window' is covering two periods of equal duration as in Figure 9.3.

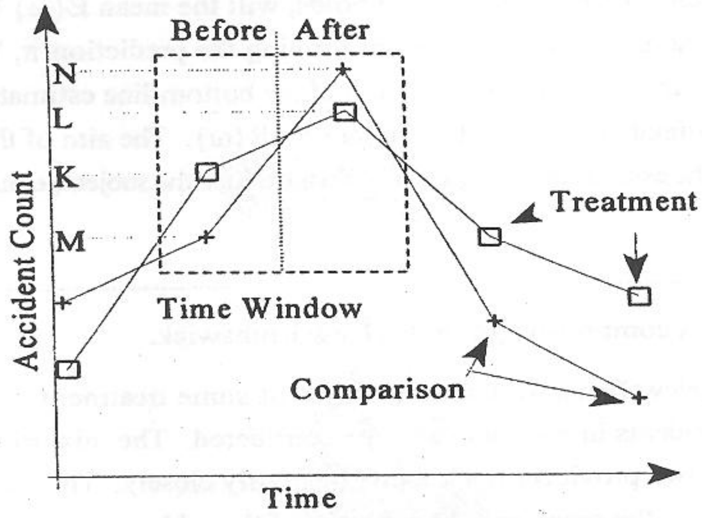


Figure 9.3. Two time series of accident counts and a 'time window'.

Over the period covered by the data, no treatment has been applied to the treatment group. The accident counts (K, L, M and N) in the time window can be organized into a 2x2 matrix as in Table 9.11.

Table 9.11. Accident counts.

	Treatment Group	Comparison Group
Before	K	M
After	L	N

A natural estimate of the odds ratio ω for one time window would be the ratio (K/L)/(M/N). However, because of the division by random variables, this would be a biased estimate. (Why this is so has been explained in connection with Equation 6.3). An approximately unbiased estimate of ω is the statistic

$$o=(KN)/(LM)/(1+1/L+1/M) \qquad \ldots 9.8$$

I will call 'o' the 'sample odds ratio'. When for a specific treatment and comparison group a time series of target accident counts is available, it is possible to calculate a sequence sample odds ratios as shown below.

Numerical Example 9.5. A comparison group for New Brunswick (continued).

In Table 9.12 I show fatal accidents counts for a short segment of the time series shown earlier in Figure 9.2. If 1971 is 'before' and 1972 is 'after', o=[(178×216)/(202×183)]/(1+1/202+1/183) = 1.03. Sliding the time window up or down, 'o's for different period pairs can be calculated. The sequence of 'o's which corresponds to the data in Figure 9.2 is shown in Figure 9.4.

Table 9.12 A segment of accident counts.

Year	New Brunswick	Saskatchewan	o
1970	183	159	
1971	178 (K)	183 (M)	1.17
1972	202 (L)	216 (N)	1.03
1973	226	199	0.82

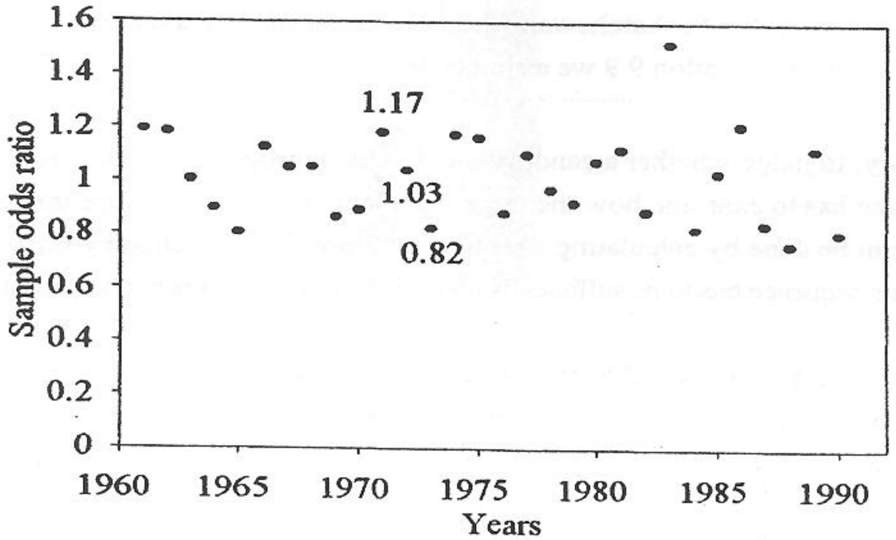


Figure 9.4. Time series of sample odds ratios for New Brunswick and Saskatchewan.

For such a sequence of sample odds ratios, one can find the sample mean denoted by $m\{o\}$, and the sample variance denoted as $s^2\{o\}$. For the data in Figure 9.2, on the basis of 30 values of o, we find $m\{o\}=1.02$ and $s^2\{o\}=0.0279$. If $m\{o\}$ is not close to 1, the comparison group is unsuitable[1] and further analysis is not necessary. If $m\{o\}$ is sufficiently close to 1, one can use the information contained in $s^2\{o\}$ to find an estimate of $VAR\{\omega\}$.

In general, the standard deviation of the sample mean is estimated by the square root of the sample variance divided by the number of data points from which it has been calculated. The sample variance here is 0.0279 and is based on 30 data points. Therefore the standard deviation of $m\{o\}$ is about $\sqrt{(0.0279/30)}=0.03$. In this case then, $m\{o\}(=1.02)$ is sufficiently close to 1 so that there is no evidence that the underlying mean ($E\{\omega\}$) differs from 1. Therefore, one can proceed to the task of estimating $VAR\{\omega\}$.

As shown in the 'derivations',

$$V\hat{A}R\{\omega\}=s^2\{o\}-(1/K+1/L+1/M+1/N) \text{ if } >0 \text{ and 0 otherwise.} \qquad ...9.9$$

Numerical Example 9.5. A comparison group for New Brunswick (concluded).

Over the thirty year period the fatal accident counts in New Brunswick (K's and L's) averaged 168/year while for Saskatchewan (the M's and N's) they averaged about 192/year. Since $s^2\{o\}= 0.0279$, by Equation 9.9 we estimate that $V\hat{A}R\{\omega\}=0.0279-2/168-2/192= 0.0056$.

In summary, to judge whether a candidate comparison group is at all suitable for a specific treatment group one has to examine how the target accidents of the two groups tracked each other over time. This can be done by calculating a sequence of sample odds ratios by Equation 9.8. The sample mean of this sequence has to be sufficiently close to 1. If not, the comparison group is unsuitable.

If a comparison group is suitable, the variance of the odds ratio can be estimated by Equation 9.9. As discussed in Section 9.1, among suitable comparison groups one should select that for which $1/N+1/M+V\hat{A}R\{\omega\}$ is smallest. Additional guidance on the choice of a comparison group has been given at the end of Section 9.2.

---

[1] As noted earlier, for a comparison group to serve, we must have $E\{\omega\}=1$. If this requirement is not fulfilled, this means that the ratio r_C is systematically larger or systematically smaller than the ratio r_T. This would render the premise of the C-G method invalid.

Derivations.

Consider a treatment group, a candidate comparison group, and a time window covering two periods. In the periods covered, **no treatment has been applied** to the treatment group. The accident counts K, L, M and N of that window and the corresponding expected values of these counts κ, λ, μ and ν can be organized into the 2x2 Table 9.13 (same as Table 9.2)

Table 9.13 Table of accident counts and of their expected values.

	Treatment Group	Comparison Group
Period 1	K, κ	M, μ
Period 2	L, λ	N, ν

1. Estimator for ω (Equation 9.8).

The odds ratio, ω, has been defined as r_C/r_T. In this, $r_C \equiv \nu/\mu$. However, since no treatment has been applied, $r_T \equiv \lambda/\kappa$ (not π/κ, as in Definition 9.2). Thus, ω can be written as the cross product of expected values

$$\omega=(\kappa\nu/\lambda\mu)$$

Expected values are never known and estimation has to be based on the available accident counts. The statistic (KN)/(LM) is a commonly used estimate of the odds ratio. However, since L and M are random variables, this is a biased estimator of ω. The approximate bias can be determined by applying Equation 6.7 with Y=(KN)/(LM). The second order partial derivatives with respect to K and N are 0. The second order partial derivatives with respect to L and M when evaluated at the expected values are $2\omega/\lambda^2$ and $2\omega/\mu^2$. When L and M are Poisson distributed, their variances are λ and μ. When Y is evaluated at the expected values of K, L, M, and N it is ω. So that $E\{(KN)/(LM)\} \approx \omega+\omega(1/\lambda+1/\mu)$. From this, $\omega=E\{(KN)/(LM)\}/(1+1/\lambda+1/\mu)$. Therefore, to remove much of the bias, use the estimator

$$o=(KN)/(LM)/(1+1/L+1/M)$$

2. Estimator for VAR{ω} (Equation 9.9).

When for a specific treatment and comparison group a time series of target accident counts is available, it is possible to calculate a sequence of sample odds ratios as illustrated in Table 9.12. This sequence has a sample variance $s^2\{o\}$. The $s^2\{o\}$ is made up of two components: the variability which is due to randomness in the accident counts, and the variance of the odds ratios - VAR{ω}. Therefore, VAR{ω} can be estimated from the difference between $s^2\{o\}$ and the variance that is due to accident counts. Were VAR{ω}=0, the variance of the sample odds ratio due to the randomness in accident counts would be

$$\mathrm{VAR}\{o|\mathrm{VAR}\{\omega\}=0\}=\omega^2(1/\kappa+1/\lambda+1/\mu+1/\nu)$$

This can again be shown by the method of statistical differentials and Equation 6.7 and is a well-known result (see, e.g., Fleiss,1981, p.63). Because we must have E{ω}=1 (and have checked that indeed m{o} is close to 1), the contribution to $s^2(o)$ of the randomness in K, L, M and N is approximately $1/\kappa+1/\lambda+1/\mu+1/\nu$. This is why we estimate

$$\hat{\mathrm{VAR}}\{\omega\}=s^2\{o\}-(1/K+1/L+1/M+1/N)$$

It can happen, especially when VAR{ω} is small, that due to the randomness in the variables on the right-hand side of the equation, the difference will be negative In this case it seems sensible to estimate $\hat{\mathrm{VAR}}\{\omega\}=0$.

9.4 A CASE STUDY: REPLACING STOP SIGNS BY YIELD SIGNS

The purpose of this section is to apply to real data the method developed and described in the preceding three sections. This should infuse into the discussion some sober realism, and help to illustrate the strengths and weaknesses of the approach, as applied. Lessons for future directions will follow.

The U.S. National Committee on Uniform Traffic Control Devices has suggested that a study be done about the safety effect of replacing STOP signs by YIELD signs. The results were to contribute to guidelines on when such a replacement of traffic control devices may be warranted. In the event, a C-G study was carried out (McGee and Blankenship, 1989). In the study, 'before' and 'after' accident counts were collected at 141 intersections where STOP signs were replaced by YIELD signs. Accident counts were collected also at 60 comparison sites. The treatment and comparison sites were in three U.S. cities: Saginaw, Pueblo and Rapid City. The conversion from STOP to YIELD signs took place in various years between 1982 and 1987. Information about traffic flow during the 'before' and 'after' periods was not used in the study; perhaps because only seldom is traffic counted at such low volume intersections and, therefore, no traffic count data was available for the 'before' periods. All police-reported accidents were considered target accidents.

The task was to estimate the safety effect of the conversion from STOP to YIELD control. Since the duration of the 'after' period was one year at some sites, two years at other sites, and only accident count totals are given, it is not possible to examine what the effect was in the first year and what it was in the second year after treatment. The assumption has to be that the safety effect of such sign conversion is the same in both years. The data are summarized in Table 9.14.

Table 9.14. Accident counts in three cities.

		Treatment Group	Comparison Group
SAGINAW	Before (2 years)	25	30
	After (2 years)	68	28
PUEBLO	Before (1 year)	4	0
	After (1 year)	12	2
RAPID CITY	Before (2 years)	12	3
	After (2 years)	26	6
TOTAL 3 CITIES	Before	41	33
	After	106	36

I will begin by the statistical analysis for the city with most accidents, Saginaw. The requisite formulae for the four-step are in Tables 9.3 and 9.4.

STEP 1. Find estimates $\hat{\lambda}$ and $\hat{\pi}$.

$\hat{\lambda}$=L=68 accidents in two years

$\hat{r}_T=\hat{r}_C=N/M/(1+1/M)=(28/30)/(1+1/30)=0.93/1.03=0.90$

$\hat{\pi}=\hat{r}_T K=0.90\times25=22.6$ accidents in two years

STEP 2. Find estimates $\hat{VAR}\{\hat{\lambda}\}$ and $\hat{VAR}\{\hat{\pi}\}$.

$\hat{VAR}\{\hat{\lambda}\}$=L=68 [accidents in two years]^2

$\hat{VAR}\{\hat{r}_T\}/r_T^2=1/M+1/N+VAR\{\omega\}=1/30+1/28+VAR\{\omega\}=0.069+VAR\{\omega\}$

$\hat{VAR}\{\hat{\pi}\}\simeq\hat{\pi}^2[1/K+\hat{VAR}\{\hat{r}_T\}/r_T^2]=22.6^2[1/25+0.069+\hat{VAR}\{\omega\}]$

$=55.7+511\times VAR\{\omega\}$ [accidents in two years]^2.

$VAR\{\omega\}$ could not be estimated from the published data. In any case, it is likely to be of secondary importance in this case, because the accident counts K, N and M are relatively small and therefore their reciprocals are large in comparison to the likely size of $VAR\{\omega\}$. If the value estimated in Numerical Example 9.5 applied here ($\hat{VAR}\{\omega\}=0.0056$), one should add only $511\times0.0056=2.9$ to 55.7.

STEP 3. Find estimates $\hat{\delta}$ and $\hat{\theta}$.

$\hat{\delta}=\hat{\pi}-\hat{\lambda}=22.6-68=-45.4$ accidents in two years or -22.7 per year.

$\hat{\theta}=(\hat{\lambda}/\hat{\pi})/[1+\hat{VAR}\{\hat{\pi}\}/\hat{\pi}^2] \approx (68/22.6)/[1+55.7/22.6^2]=2.71$

STEP 4. Find the estimates $\hat{VAR}\{\hat{\delta}\}$ and $\hat{VAR}\{\hat{\theta}\}$

$\hat{\sigma}\{\hat{\delta}\}=(\hat{VAR}\{\hat{\pi}\}+\hat{VAR}\{\hat{\lambda}\})^{1/2}=(55.7+68)^{1/2}=11.1$ accidents in two years or 5.6 per year.

$\hat{\sigma}\{\hat{\theta}\}\approx\hat{\theta}[(\hat{VAR}\{\hat{\lambda}\}/\hat{\lambda}^2)+(\hat{VAR}\{\hat{\pi}\}/\hat{\pi}^2)]^{1/2}/[1+\hat{VAR}\{\hat{\pi}\}/\hat{\pi}^2]=$

$2.71[1/68+55.7/22.6^2]^{1/2}/[1+55.7/22.6^2]=0.86.$

Thus, it appears that the STOP to YIELD conversion in Saginaw was followed by an increase of 22.7±5.6 accidents/year or a 2.71±0.86 fold increase. It is noteworthy that with relatively few accidents in the treatment and the comparison groups, such statistically firm conclusions could be reached. This may appear to contradict much of the earlier discussion in which I typically concluded that to reach statistically firm conclusions, the count of accidents has to be into many hundreds. The contradiction is only an apparent one. In fact, the result here complements (rather than contradicts) what has been noted earlier. In earlier illustrations the sought effect was a 10% to 20% change in safety. Here, we witness an increase by a few hundred percent. A standard deviation of 0.86 for a safety index of 2.71 is quite satisfactory, while the same standard deviation would make an estimated safety index of 1.1 (a 10% increase) devoid of any useful meaning. The moral is, that when the effect of some treatment on safety is large, relatively small accident counts may suffice to reach satisfactory conclusions. Conversely, when the effect on safety is modest, then, to reach reasonably precise conclusions, the accident counts need to be very large.

A comparison ratio of 0.90 has been used in this statistical analysis. A moment's reflection will show that this comparison ratio is faulty. The comparison ratio has been calculated using the total number of accidents occurring on the comparison sites in Saginaw. Its purpose was to account for changes in the safety of the treated sites between the 'before' and 'after' periods, changes that would have occurred were the STOP signs not replaced by YIELD signs. These are the changes that are due to weather, traffic, driver demography and similar factors. However, I have said at the outset, that the conversions from STOP to YIELD control took place in various years between 1982 and 1987. The 'before' and 'after' years for a conversion in, say, 1982 are entirely different from the 'before' and 'after' years for a site converted in, say, 1985. Therefore changes in weather, traffic etc. from 'before' to 'after' are also most likely different. In short, each 'before-after' pair has its specific comparison ratio. Since eight sites were converted in 1982 and 17 in 1985, the use of a single common comparison ratio common to all sites is unreasonable. Regrettably, because of the widespread but ill founded belief in the superiority of the C-G method[1], there is perhaps too little questioning of its detailed application. The use of a comparison group is for a purpose - to account for the effect of change in autonomously varying causal factors from 'before' to 'after' treatment. If this purpose is kept in mind, one would not make the mistake of using accident counts from different years on the treated and the comparison group of entities.

The Pueblo data cannot be analyzed in the same manner as the Saginaw data, because there are zero accidents on the comparison group in the 'before' period. This is another manifestation of the misunderstanding of the purpose and the potential usefulness of comparison groups in observational studies. The purpose is to correct for otherwise unaccounted for factors. If there are no major changes in traffic, the correction is likely to be relatively small. However, the error due to the randomness of accident counts which comes from using too small a comparison group, will then be large. One is attempting to remove a modest error but introducing a much larger one. If the accident counts on the comparison group of entities is small, the cure is worse than the disease.

The question of whether the accident counts of all three cities should be simply pooled will be discussed separately. The results of pooling are in the spreadsheet in Table 9.15, which is fashioned after that in Table 9.8.

Table 9.15. Statistical analysis of pooled data.

	A	B	C
1		INPUT:	
2		Treatment	Comparison
3	Before	41	33
4	After	106	36
5	$\hat{VAR}\{\omega\}$	0	
6			
7		OUTPUT:	
8	Step 1	$\hat{\lambda}=$	106
9		$\hat{r}_C=$	1.059
10		$\hat{\pi}=$	43.41
11	Step 2	$\hat{VAR}\{\hat{\lambda}\}=$	106
12		$\hat{VAR}\{\hat{\pi}\}=$	155.42
13	Step 3	$\hat{\delta}=$	-62.6
14		$\hat{\theta}=$	2.256
15	Step 4	$\hat{\sigma}\{\hat{\delta}\}=$	16.2
16		$\hat{\sigma}\{\hat{\theta}\}=$	0.632

---

[1] This belief is inherited from the tradition of conducting randomized experiments and can not be transplanted without question to the context of observational studies.

The pooled data indicate that replacing STOP signs by YIELD signs at the 141 intersections has increased accidents there by a factor of 2.25±0.63. However, for the pooled data, the earlier criticism applies with more force. Now the comparison ratio has been calculated using the 33 and 36 pooled accidents on the comparison sites. But, what does this mean and represent? Surely weather, demography, and other factors have changed one way in Pueblo (Arizona) and another in Rapid City (Michigan). Each city in each before-after period pair has a specific comparison ratio; one may not simply add up all accident counts and hope that the result gives a comparison ratio that is valid for all sites taken together.

In spite of all these quibbles and reservations, the data from all three cities consistently indicate that conversion from STOP to YIELD control was followed by a two to threefold increase of police-reported accidents. Note, however, that I did not use the word 'caused by' and preferred to say 'was followed by'. This is not mere pedantry, and I have no philosophical reservations about the notion of cause. It has been made clear earlier that in a Naive Before-After study the attribution of a change in safety to the treatment is not appropriate, because a host of other actors may have caused the same change. However, the hope was that once we account for the various other factors, either explicitly as in Chapter 8, or implicitly as in Chapter 9 (by using the comparison ratio), the attribution of change in safety to cause may be more legitimate. Why then the persistent caution?

The reason is that I suspect that the intersections that got converted from STOP to YIELD control have been **selected** partly because of their accident record. The suspicion is based on the information in Table 9.16.

Table 9.16. Accidents per intersection-year.

	Treatment Group	Comparison Group
Saginaw	0.235	0.078
Pueblo	0.057	0.000
Rapid City	0.315	0.187

Treated intersections are seen to have had more accidents prior to the conversion than those in the comparison group. This may be an indication that during the 'before' period the intersections selected for treatment had more accidents than what they would normally have. Therefore, were they left with STOP signs in the 'after' period, they would have their normal number of accidents, that is, fewer than in the 'before' period. But our statistical analysis was based on the premise that if nothing changed from 'before' to 'after' we would estimate π by K - the count of 'before' accidents.

Here, the basic premise of the statistical analysis is in danger of being wrong. One cannot compensate for an incorrect premise by rigorous statistics. Nor can one compensate for the selection bias[1] by using a comparison group, at least not simply so. How to correct for bias of this kind will be discussed in detail in Part III.

The discussion in this section is a 'reality check' of sorts. The application of the statistical machinery proved straightforward, and its output are the bottom-line estimates of δ, θ and their standard deviations. We noted that even with modest accident counts one can obtain sufficiently precise estimates when the effect of the treatment is very large. But under this solid looking surface there are cracks, even fissures. First, traffic flow data was not used. I do not believe that change in traffic can be well accounted for by using a comparison group. Second, comparison groups used were an agglomeration of accidents occurring in different periods, or even of cities. One cannot think that this use is in agreement with the two main assumptions listed early in this chapter[2] on which the C-G method rests. How to proceed when different treated entities have different comparison groups will be discussed in the next section. Third, the belief that the use of a comparison group is always good led to the use of comparison groups with so few accidents that they are virtually certain to do more damage than good. Guidance on how to choose useful comparison grops has been provided in this chapter. Fourth, there is some indication in the published data that traffic engineers choose sites to be treated for good reason, taking their past accident history into account. This invalidates the entire C-G method of prediction and renders all statistical estimates incorrect. How to cleanse estimation of this regression-to-mean bias will be discussed in Part III. In sum, if one aims for defensible results, all these issues require attention when a study is planned and carries out.

9.5 WHEN DIFFERENT ENTITIES HAVE DIFFERENT COMPARISON RATIOS

The circumstance of the case study in Section 9.4 is typical: the same treatment is applied to several entities and we wish to estimate its effect on safety. In the case reviewed, the entities (the intersections) where STOP signs were replaced by YIELD signs happened to be in various locations in three separate cities, and the signs were replaced at different times. As in Section 9.4, it is common

---

[1] The only way in which one can use a comparison group to account for a selection bias is to choose entities for the comparison group on the same criteria as for the treatment group; if entities are treated because of some aspect of their accident history, comparison entities must have the same accident history.

[2] (a) That the sundry factors that affect safety have changed from the 'before' to the 'after' period in the same manner on both the treatment and the comparison group, and (b) that this change in the sundry factors influences the safety of the treatment and the comparison group in the same way.

to think of such situations in terms of the 2×2 matrix of accident counts given in Table 9.2, and then to apply the statistical machinery pertaining to such tables. In consequence, use is being made of accident counts that are the results of pooling. Accident counts from several entities are assembled to make up the K, L, M and N in the table. However, in Section 9.4 we have recognized that entities in different cities and with diverse 'before' and 'after' periods cannot have a common comparison group. The use of a single composite comparison group with counts M and N would violate the twin assumptions on which the validity of the C-G method rests and will cast doubt on any conclusions based on it. Accordingly, the circumstance to be considered in this section is that of an observational study where a treatment has been applied to **several entities which may not have the same 'before' and 'after' periods and environment.** Therefore, each treated entity (or group of treated entities with common environments and before-after periods) has to have a separate comparison group. Typically, the number of accidents on each treated entity is small, so that one cannot reach conclusions by examining a single entity. Estimation precision would increase if the entities were pooled. But, since the comparison ratios for different 'before-after' period pairs and differing environments are likely to be different, the statistical analysis of Section 9.1 has to be modified.

This question has been examined by Tanner (1958), Griffin (1989) and Hauer (1992a). The adaptation of the four-step from Section 9.1 is straightforward. Treatment has been applied to 'n' entities (or groups of entities) labeled 1, 2, . . . , j, . . . , n. For each entity one has to obtain the estimates in Table 9.17. This table is the same as Table 9.3, except that a j has been added to identify a generic treated entity.

Table 9.17. Estimates for STEPs 1 and 2 in a 'C-G study'.

Estimates of Parameters STEP 1	Estimates of Variances STEP 2
$\hat{\lambda}(j)=L(j)$	$V\hat{A}R\{\hat{\lambda}(j)\}=L(j)$
$\hat{r}_T(j)=\hat{r}_C(j)=(N(j)/M(j))/(1+1/M(j))\approx N(j)/M(j)$	$V\hat{A}R\{\hat{r}_T(j)\}/r_T^2(j)\approx 1/M(j)+1/N(j)+V\hat{A}R\{\omega(j)\}$
$\hat{\pi}(j)=\hat{r}_T(j)K(j)$	$V\hat{A}R\{\hat{\pi}(j)\}\approx\hat{\pi}^2(j)[1/K(j)+V\hat{A}R\{\hat{r}_T(j)\}/r_T^2(j)]$

Table 9.18. STEPs 3 and 4.

$$\bar{\delta}\equiv\pi-\lambda \quad \text{...6.1}$$
$$VAR\{\hat{\bar{\delta}}\} = VAR\{\hat{\pi}\} + VAR\{\hat{\lambda}\} \quad \text{...6.2}$$
$$\bar{\theta}^* = (\lambda/\pi)/[1+VAR\{\hat{\pi}\}/\pi^2] \quad \text{...6.3}$$
$$VAR\{\hat{\bar{\theta}}\} \approx \theta^2[(VAR\{\hat{\lambda}\}/\lambda^2)+(VAR\{\hat{\pi}\}/\pi^2)]/[1+VAR\{\hat{\pi}\}/\pi^2]^2 \quad \text{...6.4}$$

As inChapter 6, (Equations 6.5 and 6.6) we define $\lambda \doteq \Sigma\lambda(j)$, $\pi \doteq \Sigma\pi(j)$, $VAR\{\hat{\lambda}\} \doteq \Sigma VAR\{\hat{\lambda}(j)\}$ and $VAR\{\hat{\pi}\} \doteq \Sigma VAR\{\hat{\pi}(j)\}$, where Σ denotes summation over all n entities. STEPs 3 and 4 are the usual equations repeated in Table 9.18. There is, however, a slight change in notation. In Chapter 6 I have noted that when it will serve clarity, $\bar{\delta}$ and $\bar{\theta}^*$ will be used instead of δ and θ^* to emphasize that what is being estimated is an 'average' safety effect for a sample of n entities.

To illustrate, a numerical example is in order. In the interest of continuity, I will analyze the data of the case study in Section 9.4. However, this is no more than an illustration because, in reality, the available data is insufficient, and 'Assumption a' introduced early in this chapter is not met. The data is insufficient for two reasons. First, one cannot use a comparison group on which zero accidents were recorded (Pueblo, Table 9.14). I will change the zero to one. Second, no data on $V\hat{A}R\{\omega(j)\}$ is available. I will assume in this example that it is 0.0055. In addition, I will have to pretend that, in each city, the 'before' and 'after' periods for the treated sites and the comparison sites are the same so that 'Assumption a' is met. As in other cases, a spreadsheet such as Table 9.19 is a convenient tool for the computations.

With the qualifications noted earlier, in this illustration we found $\hat{\bar{\theta}}=2.1\pm0.7$ and $\hat{\bar{\delta}}= -61\pm20$ accidents. The results in this section give tools for the statistical analysis of the 'average safety effect' when data are pooled from several entities and several comparison groups.

In much of present practice the comparison group is used to account for the effect of change in all factors, including traffic flow. I am of the opinion, that a comparison group and the comparison ratio should be used to account **only** for the influence of those factors that cannot be accounted for explicitly. It seems to me that, as is the habit in science, causal factors that are measured, and the influence of which is understood, are better accounted for directly. If my advice is heeded, and if the analyst uses the methods of Chapter 8 to account for the effect of, say, traffic flow, then the comparison ratio should be used subsequently to correct only for the remaining factors. This complicates matters. If, e.g., change in traffic flow on a treated entity is accounted for explicitly, one must eliminate from the comparison ratio any change that is due to traffic flow. How this might be done is discussed next.

Table 9.19. Pooling the STOP to YIELD conversion data.

	A	B	C	D	E	F
			DATA			
	Site	Treatment Before	Treatment After	Comparison Before	Comparison After	$V\hat{A}R\{\omega\}$
	j	K(j)	L(j)	M(j)	N(j)	
6	1	25	68	30	28	0.0055
7	2	4	12	1	2	0.0055
8	3	12	26	3	6	0.0055
9			**INDIVIDUAL SITES**			
10		$\hat{\lambda}(j)$	$\hat{r}_C(j)$	$\hat{\pi}(j)$	$V\hat{A}R\{\hat{\lambda}(j)\}$	$V\hat{A}R\{\hat{\pi}(j)\}$
11[1]		+C6	(E6/D6)/(1+1/D6)	+C12*B6	+C6	+D12^2*[1/B6+1/E6+1/D6+F6
12	1	68	0.90	22.58	68.00	58.41
13	2	12	1.00	4.00	12.00	28.09
14	3	26	1.50	18.00	26.00	190.78
15			**POOLED SITES**			
16	STEP	$\hat{\lambda}$	106.00	@SUM(B12..B14)		
17	1	$\hat{\pi}$	44.58	@SUM(D12..D14)		
18	STEP	$V\hat{A}R\{\hat{\lambda}\}$	106.00	@SUM(E12..E14)		
19	2	$V\hat{A}R\{\hat{\pi}\}$	277.28	@SUM(F12..F14)		
20	STEP	$\hat{\delta}$	-61.42	+C17-C16		
21	3	$\hat{\theta}$	2.09	(C16/C17)/(1+C19/C17^2)		
22	STEP	$\hat{\sigma}\{\hat{\delta}\}$	19.58	@SQRT(C18+C19)		
23	4	$\hat{\sigma}\{\hat{\theta}\}$	0.71	C21*@SQRT(C18/C16^2+C19/C17^2)/(1+C19/C17^2)		

[1] The spreadsheet formulae in row 11 are what is actually in the cells of row 12. These need to be copied into rows 13,14 and so on. The formulae in cells C16 to C23 are shown to their right.

9.6 THE MODIFIED COMPARISON RATIO

When one conducts an experiment (that is, when there is random assignment to the treatment and the 'control' group), the 2×2 matrix in Figure 9.5 is a classical archetype.

	Treatment	Control
Before		
After		

Figure 9.5. Classical archetype

In this scheme, it is the role of the control group to account for influence of change in **all** causal factors other than the treatment. In observational studies, the control group is replaced by the comparison group and its role is to account for the effect of factors other than the treatment. However, because the comparison group has not been formed by random assignment its authority is much diminished. It is now better to account directly and explicitly for change in measured and understood causal factors. Doing so modifies the role of the comparison group. It is now used to account only for the **remaining** factors, those not accounted for explicitly. Therefore, one may not estimate the comparison ratio using only the raw accident counts of the comparison group, as has been done earlier; a modification is required.

To set the stage, imagine a treated entity. The question is, as always: what would have been its expected number of target accidents during the 'after' period, had treatment not been applied? Assume that traffic flow has changed from 'before' to 'after' and that we know what the corresponding change in the expected number of target accidents is. Using the methods discussed in Chapter 8, we apply the appropriate correction to the number of 'before' target accidents. Thus, we now have an estimate of what would be the expected number of target accidents in the 'after' period had only traffic flow changed. But, additional factors (unidentified, unmeasured and not understood) have also changed from 'before' to 'after'. To account for these, there is a comparison group with accident counts M and N (see Table 9.2). In earlier sections we used M and N to estimate the comparison ratio r_C (see Equation 9.2 and Table 9.3). However, the change from M to N is not only due to the factors that remain to be accounted for, but also due to changes in traffic in the comparison group. Thus, before a comparison ratio can be estimated, the effect of changes in traffic flow in the comparison group has to be first eliminated.

In principle, this is simple to do with the tools already at hand. Suppose that during the 'before' period a road section in the comparison group had M=30 accidents and the AADT was 5000, while in the 'after' period the count of accidents was N=20 and the AADT was 4000. If on roads of this kind the safety performance function is linear, the correction for changes in traffic flow is $r_{tf}=f(A)/f(B)=4000/5000=0.8$ (see Equations 8.1 and 8.3). Thus, was the traffic of the 'after' period (A) to prevail in the 'before' period, one should expect $r_{tf}M$ accidents in the 'before' period, not M. Therefore, in estimating r_C, M should be replaced by $r_{tf}M$. Thus, in Table 9.3, use

$$\hat{r}_T=\hat{r}_{C,mod}=[N/(r_{tf}M)]/[1+1/(r_{tf}M)] \qquad \ldots 9.10$$

Consider now the case when the comparison group consists of several entities numbered 1, 2, . . . , j, . . . , m. Let M(1), M(2), . . . , M(m) and N(1), N(2), . . . , N(m) be the accident counts on these and $r_{tf}(1)$, $r_{tf}(2)$, . . . , $r_{tf}(m)$ the corresponding corrections for changes in traffic flow. In Section 9.1 the comparison ratio has been defined to be $r_C=\nu/\mu$. Correspondingly, the modified comparison ratio is defined by:

$$r_{C,mod}=\Sigma\nu(j)/\Sigma[r_{tf}(j)\mu(j)] \qquad \ldots 9.11$$

To obtain estimates of $r_{C,mod}$ and its variance is straightforward. One might even get by using the structure of equations Table 9.3, replacing N by $\Sigma N(j)$ and M by $\Sigma\hat{r}_{tf}(j)M(j)$. However, even if in principle the suggested approach is a straightforward extension of what already has been accomplished, there is a severe practical problem to overcome.

If one is to account for the changes in traffic flow directly (as in Chapter 8), data about traffic flow would have to be available for the treated entities. Were one to account for changes in traffic flow indirectly by using a comparison group (as in Chapter 9), no data about traffic flow would be needed; one only would have to live with the unlikely-to-be-true assumption that the effect of traffic change on the entities of the treatment and comparison groups was the same. But now, were one to first explicitly account for changes in traffic flow, and only then to use a comparison ratio to account for change in the remaining factors, estimates of traffic flow for both the 'before' and the 'after' period would be needed also for all entities in the comparison group. Based on previous discussion, the number of entities (accidents) in the comparison group usually has to be rather large. Therefore, 'before' and 'after' traffic flow data would be needed for the many entities of the comparison group. The discussion so far presents three imperfect options.

Option 1. As in Chapter 8 (Equation 8.2), predict $\pi=r_d r_{tf}\kappa$. In this, traffic flow and the difference in the duration of the 'before' and 'after' periods are accounted for directly. All other factors remain unaccounted for.

Option 2. As in Chapter 9 (Equation 9.5), predict $\pi = r_C \kappa$. In this the comparison ratio accounts for the difference in period durations and for the influence of all causal factors. However, it is very unlikely that the change in traffic and its effect on the safety of the entities in the treatment and comparison groups was the same.

Option 3. Predict by $\pi = r_{tf} r_{C,mod} \kappa$. In this, r_{tf} accounts directly for the effect of traffic flow and $r_{C,mod}$ for the influence of the remaining causal factors. The problem is, that information about traffic flow during the 'before' and the 'after' periods has to be available for all entities of the treatment and the comparison group.

To pursue option 3 further, I would have to undertake here the examination of the variance of the predictor $\pi = r_{tf} r_{C,mod} \kappa$. However, in view of the aforementioned practical difficulty, this seems hardly worthwhile[1]. It seems that the way out of this predicament is to consider alternate ways to predict. This task will be shouldered later, in Part III.

9.7 CHAPTER SUMMARY

An observational Before-After study with comparison groups is one of the more popular and seemingly respectable ways of estimating the effect of a treatment on safety. In this chapter I show how to do the statistical analysis for this kind of study within the framework of the 'four-step' and discuss how to choose the best of several available comparison groups. Next, I examine how the choices one makes at the study design stage influence the statistical accuracy of the results. A section is devoted to the method for assessing how well the historical record of a candidate comparison group tracks that of the treatment group. Following a case study I discuss how to do the statistical analysis when different treated entities have different comparison groups. Finally, I show how one might do the analysis when the correction for factors such as traffic flow is done first and the comparison group is used to account only for the remaining factors.

The formulae for the four-step are given in Tables 9.3 and 9.4. Table 9.8 is a sample spreadsheet for the analysis. For STEP 2, in addition to the usual accident counts K, L, M and N, an estimate of VAR{ω} is required. How to estimate it is discussed in Section 9.3. The formula to use is equation 9.9. If several candidate comparison groups are available, one should choose that for which 1/N+1/M

---

[1] It may be feasible to consider the entities of the comparison group in aggregate and correct M by an r_{tf} estimated for the group as a whole on the basis of an aggregate traffic index. Clearly, the issue is not ready to be summarized and to give advice at this time is premature.

$+\hat{VAR}\{\omega\}$ is smallest. Qualitative guidance on the choice of suitable comparison groups is given in Section 9.2

When the conduct of a C-G study is contemplated, one aims to attain a certain precision of the estimate of θ, ($VAR\{\hat{\theta}\}$). How this precision depends on the choices one can make, is specified by equation 9.7. These study-design choices pertain to the number of 'before' accidents on the treated entities (κ}, the number of 'before' accidents on the comparison entities (μ), the durations of the 'before' and 'after' period' (r_d), and the quality of the comparison group ($VAR\{\omega\}$). A spreadsheet such as that in Table 9.9 can be used to examine the sensitivity of $VAR\{\hat{\theta}\}$ to the study-design choices.

When the accidents counts K and L are the result of pooling over several treated entities, and when these treated entities can not be legitimately thought to have the same comparison group, the four-step has to be modified. The appropriate expressions are in Tables 9.17 and 9.18. Table 9.19 illustrates the analysis in spreadsheet form. If change in some causal factors is accounted for explicitly, the effect of these causal factors needs to be removed from the comparison group. In principle this is simple to do. However, severe practical problems arise as discussed in Section 9.6.

The popularity and respectability of the C-G method must have been inherited from its thoroughbred cousin - the randomized experiment. However, in most cases arising in practice, there is no random assignment to treatment. Road safety studies tend to be 'observational' in kind, and require that suitable comparison group be appointed. Therefore, the glow and glory of randomized experiments does not apply here, and a lesser image seems in order. I have tried to reduce the mystique by treating the comparison group simply as the source of information about the influence of the various unaccounted-for factors. Recall that the central task is to predict what would have been the expected number of target accidents in the 'after' period without the treatment. As a basis for this prediction we use the number of target accidents in the 'before' period. We adjust this number to account for factors measured and understood by the methods of Chapter 8. Only then do we resort to the use of the comparison group and the comparison ratio. Thus, the observational C-G study should not be thought of as a desirable prototype of study design; it is merely a way to account for the influence of factors not otherwise accounted for.

It is common to use a comparison group to account for the change in all causal factors; perhaps none have been measured, perhaps their influence on safety of is not is known. If so the results of sections 9.1-9.5 apply directly. However, it is my current belief that one should be able to estimate better and better by accounting directly and explicitly for more and more causal factors that affect safety. Therefore, the role of the comparison group should always be to account only for causal factors that are not accounted for explicitly. As our knowledge advances, the comparison group should recede in influence and use.

Endnote.

The assumption embodied in Equation 9.3 seems to be at the root of randomized experiments too, except that it does not get explicit recognition. Thus, e.g., Cochran (1983, pp. 3, 4) illustrates the essence of randomized controlled experiments as follows: "Suppose that the objective is to compare the yield per acre of a new variety of crop with a standard variety. If the new (N) and the standard (S) variety are each grown on a number of plots in the same field, the yield per plot will be found to vary from plot to plot. This variation immediately raises the problem: How far can we trust the mean difference $\bar{y}_N$-$\bar{y}_S$ over n plots of each variety as an estimate of the superiority of the new variety?"

The notation '$\bar{y}$' stands for the average yield per acre calculated from the yields of plots growing either the N or the S variety. Cochran continues: " After some false starts, this problem was finally handled roughly as follows. At least conceptually, an experiment could be so large that the difference would assume finally a fixed value, say μ_N-μ_S. The observed $\bar{y}_N$-$\bar{y}_S$ from n repetitions is regarded not as an absolute quantity, but as an estimate of the value μ_N-μ_S . . . "

Having erected the conceptual framework, Cochran is now on solid ground. He proceeds to discuss how the theory of probability can be used to describe the uncertainty surrounding the estimate of μ_N-μ_S. However, there is a step in the framework about which Cochran is not explicit. The original question is what would be the change in the yield (of a very large field) if crop variety N was used instead of S. The question was not about the expected difference of yields when S and N are sown on **different** plots. Since one cannot grow both varieties on the same field at the same time, one has to divide the field into plots, sow N in some and S in the rest. After harvest one has to **assume** that the yield of plots sown with the new variety N is indicative of **what would have been** the yield of the plots sown have been sawn with the standard variety S, had these been sown with the new variety N.

In this and many similar cases, the assumption may seem both natural, innocuous and justifiable. However, Yates (Yates, 1970, pp. 268-269) sounds a note of caution: "There is one point concerning randomization in experiments to which Fisher (R.A. Fisher, the father of randomized experimentation) always appeared to turn a blind eye. As soon as, by some appropriate random process, an experimental layout is determined, the actual layout is known . . . . Thus, in an agricultural experiment the results may indicate a fertility gradient which is very imperfectly eliminated by (a random assignment of treatments to) the blocks. Should the experimenter not then be permitted to attempt better elimination of this gradient . . . ?" Thus, randomization frees one of worry about bias only 'in the limit'; when there is a long sequence of experiments which jointly lead to a conclusion. Otherwise, whether one randomizes or not, there is always some difference between the treatment and the control group which, if not "eliminated" as Yates suggests, makes it difficult to take the validity of the assumption for granted. There is merit in the discipline of thought which directs attention to this hidden assumption and requires that it be explicitly noticed, justified if possible, or accounted for.

Why Yates thinks that Fisher turned a blind eye to the imperfections of randomization is not entirely clear since Fisher writes that: "...precision may, in appropriate cases, be much increased by the elimination of causes of variation which cannot be controlled . . . " and later that "The possibility arises from the fact that, without being equalized these differences (that is causes of variation) . . . may none the less be measured. Their estimated effects upon our final results may approximately be estimated, and the results adjusted in accordance with the estimated effects, so as to afford final precisions, in many cases, almost as great as though complete equalization had been possible" (R.A. Fisher, 1935, article 55).

It should now be clear that the Naive study (Chapter 7), the method to correct for measured and understood factors (Chapter 8) and the C-G method (Chapter 9) present three different ways for predicting 'what would have been in the 'after' period had the treatment not been applied'. All predictors used so far ($\pi=\kappa$, $\pi=r_d\kappa$, $\pi=r_d r_{tf}\kappa$, $\pi=r_C\kappa$) are similar; all seem peculiarly simple minded. One must wonder why more elaborate methods of prediction are not being exploited. An attempt to do so will make up Part III. However, before concluding the matter forming Part II, one more item needs to be addressed.

Early on, and again in Section 9.5, we have recognized that when a treatment is applied to several entities, and when δ and/or θ are estimated, these really are 'average safety effects'. Implicit in this is the recognition that the same treatment may not have the same effect on different entities. The variability of safety effect among entities is explored next.

CHAPTER 10

THE VARIABILITY OF TREATMENT EFFECT

So far we have discussed the adaptation of some commonly used 'Before-After' experimental designs to the realities of observational studies. The framework for doing so has been provided by the 'four-step' defined in Chapter 6. The 'four-step' is supposed to be a universal schema. While it proved sufficient for a unified treatment of what is present practice, it is too narrowly construed when it comes to the consideration of the variability of treatment effect.

Consider for example the case examined in Section 7.4 in which the effect of better locating signal heads and of extending the inter-green period at five intersections has been estimated. The five intersections were lumped together and treated as if they were one 'composite entity'; no use has been made of the accident counts at each intersection separately, only of their sum. We estimated there that the **average effect** of these interventions on right angle accidents was a 64%±9% reduction. The estimated effect is 'average' because the estimate pertains to the five intersections taken together. The 64% reduction does not apply to any one of these five intersections. Nor does the ±9% measure how the safety effect varies from site to site. Rather, it represents the uncertainty surrounding the estimated value of the average effect. That the effect is average implies that the true effect of the very same treatment may vary from site to site, entity to entity, and time to time. But such a variation is not part of the 'four-step' as presently formulated. In consequence, the variability of the safety effect has not, so far, been estimated within the framework provided. A broader formulation of the 'four-step' is needed.

The point of departure in this chapter is the recognition that when the same treatment is applied on different occasions and sites, its true safety effect may vary from application to application. If these occasions of application are thought of as 'trials', and the treatment effect for each occasion of application is considered an 'outcome', then one can speak of the **mean effect** in a set of treatment applications, and also of the **variance of the effect** in a set of treatment applications. This variance may be large or small. Should it turn out to be small, the treatment may be expected to have approximately the same effect in other (future) applications. Should the variance be large, the effect of the treatment for another application is difficult to predict; it may be harmful in some instances and very useful in others. What then is the site-to-site, entity-to-entity or time-to-time variance of a treatment? How can it be estimated? What is the reason for our interest in this variance? These and similar questions require clarification.

There are two main reasons for which we may be interested in the safety effect of a treatment. First, we may wish to know what the effect of a treatment was for a specific set of treated entities. If the treatment was a success, the boss or the public will want to know; if it was harmful, perhaps the treatment should be modified or removed. Many think that this kind of 'safety effect estimation' is an essential part of sound safety management, a view about which I have reservations[1]. Second, one may not be interested in the specific entities that were treated, except inasmuch as they provide clues to the safety effect of the treatment in general. One hopes that such clues find their way into the professional lore as articles, books or courses, and eventually are incorporated in standards, warrants and other tools that affect professional practice. Thus, the second reason for our interest in the estimation of the safety effect of treatments is that these estimates are a guide to action in the future, often elsewhere, and by others. This, I think, is the more compelling reason for wanting to estimate the effect of a treatment. As knowledge about the effect of a treatment accumulates, fact-based decision-making becomes both possible and better.

To make cost-effective decisions about the implementation of treatments or about safety-related standards or warrants, it is usually sufficient to have an estimate of the average effect. The need to estimate the variance of the safety effect requires a separate justification. The estimate of the variance of the safety effect of a treatment is needed for two main purposes:

a. In deciding whether to implement some treatment, one may wish to know, say, what is the chance that the treatment will reduce safety.

b. The existence of variability in safety effect is evidence that the same treatment is more effective for some entities or circumstances than for others. To manage safety well, it is important to find what are the traits of entities and the circumstances that make a treatment more effective, and which are those traits or circumstances in which a reduction to safety may follow.

Item 'b' is a key to progress in road safety management. If one inquires about the reasons why the same treatment is more effective in one case than in another, there is a chance to obtain an ever more detailed understanding of cause and effect and, as a result, to manage safety better. Thus the point of departure for a more general framework for the analysis and interpretation of Before-After studies, is the recognition that the effect of a certain treatment may vary from entity to entity. As a result, the broader **aim is to estimate both the mean and the variance** of the safety effect,

---

[1] Seldom is this kind of estimation possible and rarely is it done professionally. What results are obtained and whether they are publicized is likely to be more influenced by the interests of the institution than of the road users.

either in a specific set of entities to which the treatment has been applied, or in a hypothetical large population of entities. This requires an expansion of the conceptual and notational arsenal. The basic building blocks devised in Chapter 6 need enriching.

10.1 THE EXPANDED 'FOUR-STEP'

Consider the following problem: The same intervention is applied in several instances (e.g., to several entities) and its effect is estimated in each case. The effect of the intervention may vary from one instance of application to another. If we knew exactly what the effect in each instance of application was, we could calculate the sample mean and variance of the effects. However, the effect is never known exactly. For each instance of application we can only have an **estimate** of the effect and also an estimate of the variance of the estimated effect. The question is how, from this information, can one learn something about the variance of effects among treatment applications.

Imagine that the treated entities are labeled 1, 2, . . . , j, . . . , n and the corresponding indices of effectiveness of the treatment for these entities are denoted by $\theta(1), \theta(2), \ldots, \theta(j), \ldots, \theta(n)$. The overall usefulness of the treatment for these entities is judged by the sample mean of these; this has been called the average effect, $\bar{\theta} = \Sigma\theta(j)/n$. How $\bar{\theta}$ is estimated has been discussed in Section 9.5. However, to say what the effect of this treatment is expected to be for an entity not yet treated, one has to construe the evidence from the already treated entities, as a clue to something more general. For that purpose, imagine a population of entities for which the 'n' treated entities can be considered a random (or representative) sample. The mean of the θ's in this imagined population is denoted as $E\{\theta\}$, their variance as $VAR\{\theta\}$. To anticipate the future safety effect of the treatment for an entity of this population, we require estimates of $E\{\theta\}$ and of $VAR\{\theta\}$.

So far, the 'four-step' procedure has been confined to the estimation of $\bar{\theta}$. Because the treated entities are considered a random sample from the population, the estimate of $\bar{\theta}$ is also an estimate of the population mean $E\{\theta\}$. Now the 'four-step' needs to be expanded to entail the estimation of $VAR\{\theta\}$. The estimates $\hat{\theta}(j)$ and $\widehat{VAR}\{\hat{\theta}(j)\}$ will serve as raw material. The procedure is straightforward:

1. Apply the single-entity 'four-step' of Chapter 6 to obtain $\hat{\theta}(j)$ and $\widehat{VAR}\{\hat{\theta}(j)\}$ for entities $j=1, 2, \ldots, n$.

2. Using the estimates $\hat{\theta}(j)$ compute the sample variance $s^2\{\hat{\theta}\}=\Sigma[\theta(j)-\bar{\theta}]^2/(n-1)$.

3. Using the estimates $V\hat{A}R\{\hat{\theta}(j)\}$, j=1, 2, . . . , n, compute their average $avg(\hat{V}) = \Sigma V\hat{A}R\{\hat{\theta}(j)\}/n$.

4. As shown in the 'derivations' at the end of this section, $VAR\{\theta\}$ can now be estimated by:

$$V\hat{A}R\{\theta\}=s^2(\hat{\theta}) - avg(\hat{V}) \qquad \ldots 10.1$$

That is, the variance of θ in the population of treatments is estimated by computing the sample variance of the estimates of indices of safety effect, and subtracting from this the average of the variances of these estimates.

The expanded 'four-step' now takes the following form:

The Expanded 'Four-Step'.

STEP 1.	**For j=1, . . ., n estimate $\lambda(j)$ and $\pi(j)$.**
STEP 2.	**For j=1, . . . , n estimate $VAR\{\hat{\lambda}(j)\}$ and $VAR\{\hat{\pi}(j)\}$.**
COMPOSITE ENTITY STEP.	**Estimate λ, π, $VAR\{\hat{\lambda}\}$, and $VAR\{\hat{\pi}\}$ using Equations 6.5 and 6.6.**
STEP 3.	**Estimate $\bar{\delta}$ and $\bar{\theta}$.**
STEP 4.	**Estimate $VAR\{\hat{\bar{\delta}}\}$ and $VAR\{\hat{\bar{\theta}}\}$.**
TREATMENT VARIANCE STEP	**Estimate $VAR\{\theta\}$ using Equation 10.1.**

Derivations.

The argument leading to Equation 10.1 could be made relatively simply by decomposing the total variance into one component that is due to the variation of θ's between entities and another component that is due to the variance with which the θ's are estimated. This route has been followed, e.g., by Cochran (1954) in his classic paper: 'The combination of estimates from different experiments'. I will prefer a more long-winded derivation because it is more explicit, gives clear meaning to all terms, and does not require any assumptions.

Analysis.

The results here are perhaps of more general interest than to road safety. In particular, they apply to the basic tasks of statistical meta-analysis. For this reason, and to avoid notational complexity, I will use a general notation, one that is unrelated to that used in this book so far. The linkup with the safety index estimation will be in the next subsection.

Consider a discrete random variable X that can take on k values $x_1, x_2, \ldots, x_i, \ldots, x_k$. Assume that there exist n distributions of X with means and variances $(m_1, v_1), (m_2, v_2), \ldots, (m_j, v_j), \ldots, (m_n, v_n)$. Denote by $p_j(x_i)$ the probability that in the j-th distribution $X=x_i$. A 'trial' consists of taking just one observation of X from each of the n distributions. This is the defining feature of the problem at hand. If the trial was to be repeated many times, the relative frequency with which the value x_i is observed would approach

$$P(X=x_i)=[p_1(x_i)+p_2(x_i)+\ldots+p_n(x_i)]/n. \qquad \ldots a$$

This completely describes the probability distribution of X for a long sequence of such trials.

We are interested in the mean E{X} and variance VAR{X} of the probability distribution given by (a). Specifically we wish to express E{X} and VAR{X} in terms of the $(m_1, v_1), (m_2, v_2), \ldots, (m_j, v_j), \ldots, (m_n, v_n)$. To this end we write:

$$E\{X\} \doteq \sum_{i=1}^{i=k} x_i\, P(X\ \ x_i)\ \ \frac{\sum_{i=1}^{i=k} x_i [p_1(x_i)\ \ p_2(x_i)\ \ldots\ p_n(x_i)]}{n}\ \ \frac{\sum_{j=1}^{j=n} m_j}{n} \doteq \bar{m} \qquad \ldots b$$

$$E\{X^2\} \doteq \sum_{i=1}^{i=k} x_i^2\, P(X\ \ x_i)\ \ \frac{\sum_{i=1}^{i=k} x_i^2 [p_1(x_i)\ \ p_2(x_i)\ \ldots\ p_n(x_i)]}{n}\ \ \frac{\sum_{j=1}^{j=n} (v_j\ \ m_j^2)}{n} \qquad \ldots c$$

Since VAR{X}=E{X^2}-E^2{X}, and using the result in 'b' and 'c',

$$VAR\{X\}\ \ \frac{\sum_{j=1}^{j=n} (v_j\ \ m_j^2)}{n}\ \ \frac{n\bar{m}^2}{n}\ \ \frac{\sum_{j=1}^{j=n} v_j}{n}\ \ \frac{(\sum_{j=1}^{j=n} m_j^2)\ \ n\bar{m}^2}{n} \doteq \bar{v}\ \ s_m^2 \qquad \ldots d$$

In this $\bar{v}$ is the sample mean of the variances $v_1, v_2, \ldots, v_j, \ldots, v_n$ and s^2_m is the (biased[1] form of the) sample variance of $m_1, m_2, \ldots, m_j, \ldots m_n$. It follows that

$$s^2_m = VAR\{X\} - \bar{v} \qquad \ldots e$$

For estimation, we use the n observed values $x_1, x_2, \ldots, x_n$ of one trial and estimates of their variances $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_n$. From the $x_1, x_2, \ldots, x_n$ we compute the sample variance s^2_x which is then used to estimate VAR{X} in Equation 'e'. From the $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_n$ we compute their average denoted by avg($\hat{v}$) which is then used to estimate of $\bar{v}$ in Equation 'e'. Therefore, the difference of the statistics s^2_x - avg($\hat{v}$) is an estimate of s^2_m.

So far, the frame of reference were trials in which one observation was taken from each of n distributions. The next step is to imagine that there exist very many such distributions, a population of distributions, and that the n distributions considered so far are a random (representative) sample from this population. In this population of distributions, the m's have a mean E{M} and a variance VAR{M}. As usual, the sample variance of the m's, the s^2_m, is an estimate of the VAR{M}. Therefore, for that population for which our sample is a random one, we may estimate:

$$V\hat{A}R\{M\} = s^2_x - avg(\hat{v}) \qquad \ldots f$$

This concludes the analysis. Although the derivation is for a discrete random variable X which can take on a finite number of values, a similar line of reasoning would lead to the same result if X was taken to be a continuous random variable. While in the derivation we used probability mass functions $p_j(x_i)$, specifying what they are was not necessary. Therefore, Equation 'e' holds irrespective of what the $p_j(x_i)$ are, as long as the setting corresponds to the specified 'trial'; namely that one takes one observation of X from each of n distributions, and each distribution has its own mean and variance.

<u>Interpretation for road safety.</u>

We now describe the correspondence between the results obtained for the general case, in which one random variable is obtained from each of n distributions, and the specific problem of estimating the variance of the safety index θ that is of interest here. In our case, the same intervention has been applied in n instances. In each instance the intervention has had an (unknown) 'true safety index' denoted by θ(j), j=1, . . . , n; thus the θ(j) correspond to the m_j in the analysis above. We know how to obtain estimates $\hat{\theta}$(j) of the θ(j); these $\hat{\theta}$(j) correspond to the x_j in the

---

[1] Usually the sample variance is defined as having n-1 in the denominator, not n.

analysis above. Finally, we also know how to estimate the variance of $\hat{\theta}(j)$; these $\hat{VAR}\{\hat{\theta}(j)\}$ correspond to the $\hat{v}_j$ in the analysis.

Let now $s^2(\hat{\theta})$ denote the sample variance computed from the $\hat{\theta}(1), \hat{\theta}(2), \ldots, \hat{\theta}(n)$. Because the $\hat{\theta}(j)$ correspond to the x_j, $s^2(\hat{\theta})$ corresponds to s^2_x in Equation 'f'. Let avg($\hat{V}$) denote the sample mean of the variances computed from the $\hat{VAR}\{\hat{\theta}(1)\}, \hat{VAR}\{\hat{\theta}(2)\}, \ldots, \hat{VAR}\{\hat{\theta}(n)\}$. Because the $\hat{VAR}\{\hat{\theta}(j)\}$ correspond to the $\hat{v}_j$ in the analysis, avg($\hat{V}$) corresponds to avg($\hat{v}$) in Equation 'f'. If so, Equation 'f' interpreted for the present case is $\hat{VAR}\{\theta\}=s^2(\hat{\theta})$ - avg($\hat{V}$). This is Equation 10.1.

10.2 AN ILLUSTRATION: RAISED PAVEMENT MARKERS

Griffin (1990 b) provides 'daytime' and 'nighttime' accident counts for two years before and two years after the installation of raised pavement markers on 86 road sections in Texas. The first two and last two of the 86 rows are shown in Table 10.1. Griffin assumes that the installation of raised pavement markers affects only nighttime accidents and that daytime accidents may be used for 'comparison'[1].

Table 10.1. Road section data.

Road Section	Night (Treatment)		Day (Comparison)	
	Before K(j)	After L(j)	Before M(j)	After N(j)
1	27	23	101	97
2	9	7	8	8
...				
...				
85	5	11	9	7
86	7	7	10	8

With this as data, one can obtain the estimates in Table 10.2.

[1] Whether this is a justifiable assumption is not easy to establish.

Table 10.2. Computations.

1	2	3	4	5	6	7	8
Site-j	$\hat{\lambda}(j)$	$\hat{r}_C(j)$	$\hat{\pi}(j)$	$V\hat{A}R\{\hat{\lambda}(j)\}$	$V\hat{A}R\{\hat{\pi}(j)\}$	$\hat{\theta}(j)$	$V\hat{A}R\{\hat{\theta}(j)\}$
1	23	0.95	25.68	23	37.74	0.847	0.065
2	7	0.89	8.00	7	23.11	0.643	0.112
...							
...							
85	11	0.70	3.50	11	5.56	2.162	1.204
86	7	0.73	5.09	7	9.53	1.005	0.276
	1471 $\hat{\lambda}$		1250.33 $\hat{\pi}$	1471 $V\hat{A}R\{\hat{\lambda}\}$	3133.60 $V\hat{A}R\{\hat{\pi}\}$	0.621 $s^2\{\hat{\theta}(j)\}$	0.514 $avg[V\hat{A}R\{\hat{\theta}(j)\}]$

Thus, by STEP 1 in Table 9.16, $\hat{\lambda}(1)=23$, $\hat{r}_C(1)=(97/101)/(1+1/101)=0.951$, $\hat{\pi}(1)= 0.95\times27 =25.68$. By STEP 2 of Table 9.16, $V\hat{A}R\{\hat{\lambda}(1)\}=23$ and $V\hat{A}R\{\hat{\pi}(1)\}=25.68^2\times[1/27+1/101+1/97] =37.74$. This is based on the assumption that VAR{ω}=0. Since, in this case, it is the daytime accidents on the same road sections which serve as comparison accidents, the assumption may be justifiable. Using now Equations 6.3 and 6.4, $\hat{\theta}(1)=(23/25.68)/(1+37.74/25.68^2)=0.847$ and $V\hat{A}R\{\hat{\theta}(1)\}= 0.847^2\times(23/23^2+37.74/25.68^2)/(1+37.74/25.68^2)^2=0.065$. These results are tabulated in the row of site 1 in Table 10.2. As before, only the rows for the first and last two sites are shown.

The values of $\hat{\lambda}$, $\hat{\pi}$, $V\hat{A}R\{\hat{\lambda}\}$ and $V\hat{A}R\{\hat{\pi}\}$ in the last row are as defined in Chapter 6 by Equations 6.5 and 6.6. Now, using Equations 6.3 and 6.4 from Table 9.18) we find:

$$\hat{\theta} = (1471/1250.33)/\ (1+3133.60/1250.33^2)=1.174 \text{ and}$$
$$V\hat{A}R\{\hat{\theta}\}=1.174^2\times(1471/1471^2+3133.60/1250.33^2)/\ (1+3133.60/1250.33^2)^2 =0.0037$$

The average safety effect of raised pavement markers is estimated to be an increase of nighttime accidents by 17.4%. The standard deviation of this estimate of average safety effect is $\sqrt{0.0037}=0.061$ or ±6.1%.

The sample variance of the 86 values of $\hat{\theta}(j)$ in column 7 is 0.621 and the average of the $V\hat{A}R\{\hat{\theta}(j)\}$ in column 8 is 0.514. By Equation 10.1,

$$V\hat{A}R(\theta)=0.621-0.514 = 0.107 \text{ or } \hat{\sigma}\{\theta\}= \sqrt{0.107} = 0.328.$$

Thus, the effect of installing raised pavement markers on nighttime accidents can be quite diverse. While in an average application one may expect a 17.6% increase in nighttime accidents, one standard deviation on either side makes for a range from a 50% increase to a 15% decrease in nighttime accidents. If one can approximate the distribution of θ's by a Gamma distribution with a mean of 1.176 and variance of 0.107, it would look like that in Figure 10.1. Thus, in about a third of road sections one may expect that raised pavement markers are expected to cause a decrease in nighttime accidents, and in two-thirds to cause an increase. Naturally, the matter should not rest there. One would like to know whether one can ascertain the characteristics of those road sections where this treatment is beneficial and on those where it is detrimental to safety.

Present practice is to estimate only the average safety effect and, in some cases, the precision with which this average is estimated. For most cost-effect considerations this is sufficient. However, for the two reasons given earlier it may be important to know how variable is the effect of the same treatment, when applied to different entities and on different occasions. Equation 10.1 provides the necessary tool to estimate the variance of the index of safety effect. It should be of particular interest when results of published studies are subject to a Meta-Analysis. A brief review of Meta-Analysis is given next.

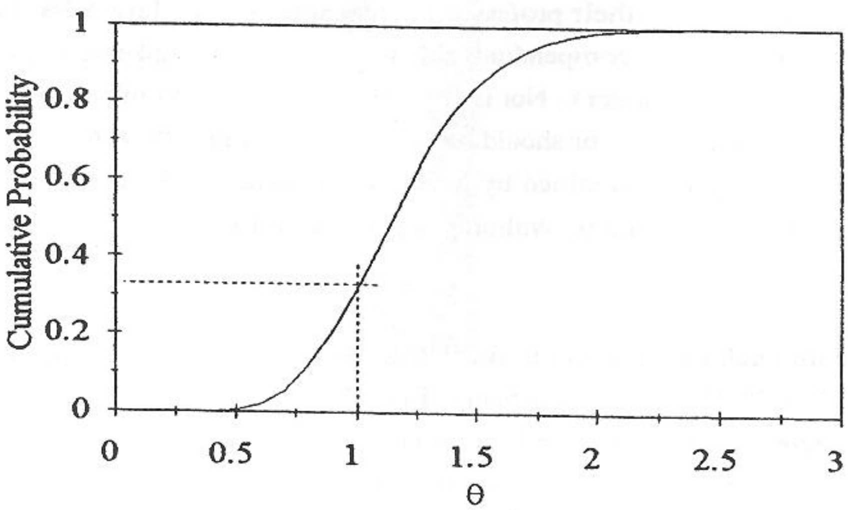


Figure 10.1. The distribution of θ's for applications or raised pavement markers.

10.3 APPLICATION TO META ANALYSIS

The main reason for asking what effect a treatment had on the safety of some entities is to have factual guidance for the future. Every such inquiry adds to the edifice of professional knowledge and to the quality of professional advice. Almost never will a single study settle a question once for all. The usual state of affairs is that over time several studies about the effect of the same treatment are published. The results, taken together, represent 'what is known' at some point in time. Given that the accumulated results are usually diverse, one may ask: how should they be 'taken together'? Are there some formal means by which the collection of separate studies can be made to coalesce into the edifice of fact-based professional lore?

The question of how to extract professional knowledge from published accounts is not specific to road safety. However, in this field, it seems particularly troubling. In some other spheres of professional endeavor, accumulated evidence from research is periodically and critically reviewed and published. The professional only needs to look up the latest compendium to find out what is known to-date about the effect of this or that treatment, or what is the prognosis for this or that disorder. The highway design engineer, the transportation planner, the traffic engineer, or the urban planner, cannot reach to their professional bookshelf to find how what they do will affect safety. There is no authoritative compendium about the safety consequences of design and planning decisions to guide these practitioners. Nor is it practical for them to embark on a literature review every time a safety question arises or should be considered. Even if they did, they would encounter a bewildering array of results obtained by a variety of means. The upshot is that professional decisions often need to be made without fact-based information about their likely safety repercussions.

A few compendia attempting to fulfil this need do exist. In North America there is the 'Synthesis' (FHWA, 1982) and its descendant (FHWA, 1992). These identify informative studies on a range of selected subjects and briefly summarize the findings of each[1]. An earlier publication (FHWA, 1981) does contain numerical estimates of 'accident reduction factors'. These are widely deemed to be ill founded because of flaws in method and data. Similar doubts attach to the short list of 'accident reduction factors' included in the 'Toolbox' (ITE, 1993). In a small scale effort Persaud (1992) reviewed and summarized what is known on a range of measures and treatments. A more concerted effort has been made in the Nordic countries (Statens Vägverk, 1981, Elvik et al., 1989, Johannssen, 1982). The book by Ogden (1996) comes perhaps closest to an international

[1] It is perhaps telling that the authors of the later version (FHWA, 1992) found it necessary to omit many studies included in the earlier version (Synthesis, FHWA, 1982) as being either too old or too questionable and did not find much new material to take their place.

review of what is known about the safety effect of traffic engineering and highway design features.

Publications of this kind are usually the product of an informal process. An expert, or a group of experts, assemble and review the published literature, and then apply judgement to extract its lessons. Because the constituent studies vary in quality, in the amount of data used, in methods of analysis, in the validity of inferences, and in how results are presented, the task of judging, reconciling and summarizing their results is a subjective one. This raises an intriguing question. Each study about the safety effect of a treatment uses statistical methods to extract estimates from its data. Results of each study are presented in statistical terms: as point estimates, as confidence intervals, or as tests of significance. Why then must one summarize the collective results of several studies by subjective, non-quantitative and non-statistical means? Is it not possible to combine separate quantitative results by a formal quantitative procedure?

Much thought has been given to this question in agriculture, psychology, education, medicine and sociology. In these disciplines, a quantitative alternative, or perhaps a supplement, to the literature-review-expert-judgement has emerged. The quantitative method for the combination of results from different studies has been given[1] the name **Meta Analysis.** Some books on this subject are listed at the end of this section. Recently meta-analytic methods have been applied in road safety (see, e.g., Elvik, 1994, 1995a, b).

Chapters 6, 7, 8 and 9 dealt with the question of how to obtain estimates of safety effect when a treatment has been applied to a single entity or to a composite entity. In this chapter I have shown that if a treatment is applied to several entities, it is possible to estimate not only the average effect of the treatment but also its variance among entities[2]. From here it is only a small conceptual step to draw the parallel between a set of entities to which a treatment has been applied and a set of studies (or composite entities) to which the treatment has been applied (figure 10.2).

---

[1] The term Meta Analysis is attributed to Glass (1976). However, as early as 1904 Karl Pearson used five data sets of data to summarize the relation between immunity from infection and inoculation.

[2] See the endnote in Chapter 6 for a discussion of the limitations of θ (the index of effectiveness) as a measure of safety effect.

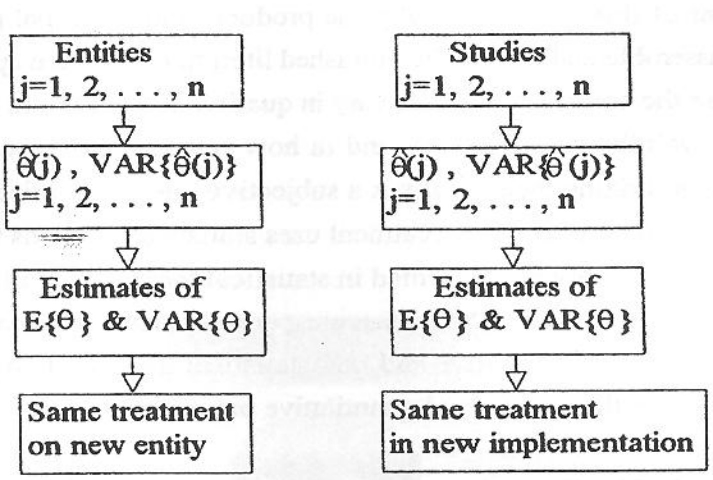


Figure 10.2. Similarity in procedure.

When for each of the treated entities 1, 2, . . . , j, . . . , n we had the estimates $\hat{\theta}(j)$ and $\hat{VAR}\{\hat{\theta}(j)\}$, I have shown (Section 10.1) how to estimate the mean ($E\{\theta\}$) and the variance ($VAR\{\theta\}$) of the index of effectiveness θ. These two estimates give guidance on what one may expect if the same treatment is to be applied to a new entity (provided that it can be deemed to be from a population of entities of which the n treated entities are a random sample). In parallel, let there be studies 1, 2, . . . , j, . . . n, each of the same treatment. Thus, we have estimates $\hat{\theta}(j)$ and $\hat{VAR}\{\hat{\theta}(j)\}$, one for each study. Using now the very same procedure, one can estimate $E\{\theta\}$ and $VAR\{\theta\}$. Now these two quantities describe what one may expect if the same treatment is applied again in a new circumstance (provided that the new circumstance may be considered to be from a population for which the set of n studies is a random sample.)

To illustrate, I will use data from Elvik's (1995b) ground breaking paper describing a meta-analysis of the safety effect of road lighting. His data come from 37 separate studies conducted in different countries at various times. Each study contains one or more quantitative 'results'. Thus, e.g., a study may have two separate results, one for rural and one for urban roads. In all Elvik assembled 142 results. The target accidents were those occurring at night. Various study types were used. Some studies compared nighttime accidents before and after illumination (Before-After studies). Other studies compared roads with and without illumination (Cross-Section studies). Sometimes the comparison accidents were nighttime accidents on roads that remained unlit; sometimes daytime accidents on the treated road sections were used for comparison. In some studies a combination of both kinds of comparison groups was used. The circumstances differed from result to result. The environment may have been urban, rural or motorway; accidents may have been to vehicles or pedestrians, at junctions or between them; the recorded accidents may have been fatal, non-fatal injury or property damage only etc. The data in Table 10.3 are a small subset extracted from Elvik's Table 4. Given are pedestrian injury accident counts for 13 'results' of illumination in an urban environment. The studies are all of the C-G type with daytime accidents as the comparison group. The computations for each 'result' in Table 10.4 are exactly the same as the computations for each 'entity' in Table 10.2

Table 10.3. Data for 13 studies of illumination in urban areas.

Result	Author	Year	Country	Severity	K	L	M	N
1	Tanner	1955	UK	Fatal	6	1	1	4
2	Tanner	1955	UK	Injury	31	19	73	71
3	Tanner	1958	UK	Injury	144	85	314	323
4	Christie	1966	UK	Injury	7	0.5	1	1
5	Fisher	1971	Austria	Injury	10	6	23	18
6	Fisher	1971	Austria	Injury	16	10	18	19
7	Fisher	1971	Austria	Injury	15	6	16	20
8	Fisher	1971	Austria	Injury	17	6	28	20
9	?	?	USA	Injury	175	122	221	294
10	?	?	USA	Fatal	84	22	42	46
11	Pegrum	1972	Austria	Injury	32	13	57	58
12	Fisher	1977	Austria	Injury	162	87	219	276
13	Schwab	1982	Denmark	Injury	51	19	44	51
Sums					750	396.5	1057	1201

The values of $\hat{\lambda}$, $\hat{\pi}$, $V\hat{A}R\{\hat{\lambda}\}$ and $V\hat{A}R\{\hat{\pi}\}$ in the last row of Table 10.4 are as defined in Chapter 6 by Equations 6.5 and 6.6. Using now Equations 6.3 and 6.4 from Table 9.17 we find:

$$\hat{\theta} = (396.5/860.4)/\ (1+2766.6/860.4^2)=0.46 \text{ and}$$

$$V\hat{A}R\{\hat{\theta}\}=0.459^2\times(396.5/396.5^2+2766.6/860.4^2)/\ (1+2766.6/860.4^2)^2 =0.00131$$

Thus, illumination in urban areas seems to reduce pedestrian injury accidents on the average by 54%[=(1-0.46)×100]. The standard deviation of this average is $\sqrt{0.00131}=0.036$ or ±3.6%. The sample variance of the 13 values of $\hat{\theta}(j)$ is 0.0402 and the average of the $V\hat{A}R\{\hat{\theta}(j)\}$ is 0.0251. By Equation 10.1

$$V\hat{A}R(\theta)=0.0402-0.0251=0.0151 \quad \text{or} \quad \hat{\sigma}\{\theta\}=\sqrt{0.0151}=0.12.$$

Table 10.4. Computations.

result - j	$\lambda(j)$	$r_c(j)$	$\hat{\pi}(j)$	$V\hat{A}R\{\hat{\lambda}(j)\}$	$V\hat{A}R\{\hat{\pi}(j)\}$	$\hat{\theta}(j)$	$V\hat{A}R\{\hat{\theta}(j)\}$
1	1	2.000	12.0	1.0	204.0	0.034	0.000
2	19	0.959	29.7	19.0	53.1	0.603	0.036
3	85	1.025	147.7	85.0	288.3	0.568	0.008
4	0.5	0.500	3.5	0.5	26.3	0.045	0.001
5	6	0.750	7.5	6.0	11.2	0.667	0.113
6	10	1.000	16.0	10.0	43.7	0.534	0.056
7	6	1.176	17.6	6.0	55.8	0.288	0.021
8	6	0.690	11.7	6.0	19.9	0.447	0.047
9	122	1.324	231.8	122.0	732.6	0.519	0.006
10	22	1.070	89.9	22.0	463.9	0.232	0.005
11	13	1.000	32.0	13.0	67.6	0.381	0.018
12	87	1.255	203.2	87.0	593.2	0.422	0.004
13	19	1.133	57.8	19.0	206.9	0.310	0.010
	396.5 $\hat{\lambda}$		860.4 $\hat{\pi}$	396.5 $V\hat{A}R\{\hat{\lambda}\}$	2766.6 $V\hat{A}R\{\hat{\pi}\}$	0.0402 $s^2\{\hat{\theta}(j)\}$	0.0251 $avg[V\hat{A}R\{\hat{\theta}(j)\}]$

In this illustrative example, the average experience of these thirteen results is that illumination in urban areas reduces pedestrian accidents by a factor of 0.46. This experience varies somewhat from time to time and place to place. Using the rule of thumb mentioned in Section 7.3, one might expect some 65% of the cases in which an urban area is illumined to have θ within the 0.34 - 0.58 range. If one can approximate the distribution of θ's by a Gamma distribution with a mean of 0.46 and a variance of 0.0151 it would look like that in Figure 10.3. In this figure, the 0.34 - 0.58 range of θ's corresponds approximately to the 20th and 85th percentiles.

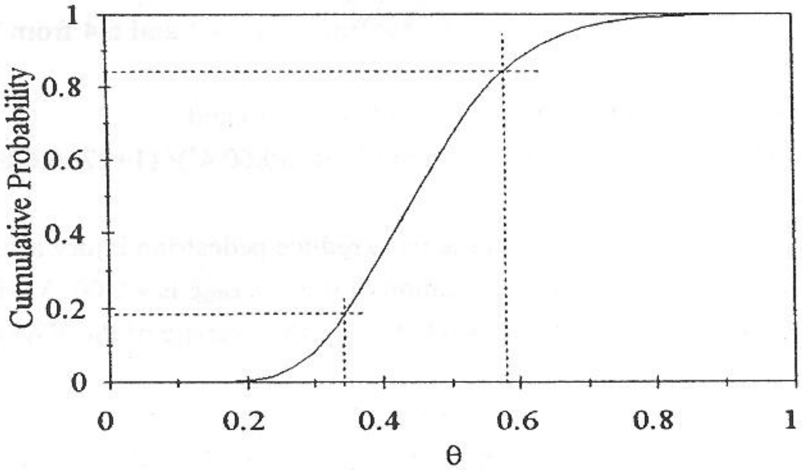


Figure 10.3. The distribution of θ's for pedestrian accident in applications of illumination in urban areas.

A note of caution is in order. The predictions $\hat{\pi}(j)$ in Table 10.4 are based on the assumption that daytime accidents on a road section are an ideal comparison group for nighttime accidents which are the target of treatment. That is, I have assumed that $VAR\{\omega(j)\}=0$. Since this is not likely to be true, the estimates $V\hat{A}R\{\hat{\pi}(j)\}$ in Table 10.4 are all somewhat lower then what they should be. If so, the estimates $V\hat{A}R\{\hat{\theta}(j)\}$ in the same table are also too small. It follows that in Equation 10.1 we are subtracting too small a number. Therefore, our estimate of $VAR\{\theta\}$ (=0.0151) is likely to be too large. To obtain a better estimate we would need to know how well daytime and nighttime accidents track each other over the years. This research is at present undone.

In summary, if we have the estimates $\hat{\theta}$ and $V\hat{A}R\{\hat{\theta}\}$ for several instances of treatment application, one can estimate how the effect of the treatment varies from one application to another. We may consider the fitting of one road section with raised pavement markers to be an 'instance of application' as has been done in Section 10.2. We may also consider a published account of a study about the effect on pedestrian accidents of illuminating a certain urban area to pertain to an 'instance of application', as in this section. The analysis of the variability of safety effect is essentially identical in both cases.

I should not leave the impression that there is not more to meta-analysis than computation. Nor should one forget three important biases involving the raw material for any meta-analysis. First, a decision was made **whether to study** the effect a treatment in a certain case of its application. This decision may have been influenced by the interest of those who favored or opposed the implementation of the treatment, or who may have had an inkling of what the result of a study could be, or who had little to gain from a positive result but a lot to loose from a negative one. What is now available for meta-analysis is determined by such decisions. Second, the decision **whether to publish** the results of a study may be influenced by what has been found. There may be more incentive to publish accounts of successes than of failures. This decision also affects what raw material is available. Third, there is a measure of control over **what to write** in the conclusions. This comes through the myriad of untraceable choices that need to be made during data collection and analysis, and which influence the conclusions in foreseeable directions. Some such directions may be deemed pleasing to the analyst or to interested parties. The first two biases, if substantial, make the sample of studies available for a meta-analysis to be non-representative. Since such a sample does not represent the population of potential treatment applications, one cannot use the results of a meta-analysis to make predictions about what might be the effect of this treatment in the next application.

This agnostic argument can be countered as follows. Every one of the three potential biases (whether a study is done, whether its results are published, and in which direction its conclusions are steered) applies to single, individual, studies. Even so, it is normal practice to use quantitative statistical methods of data analysis to reach and state conclusions in each individual study. Indeed, it would be peculiar nowadays to suggest that it is better to interpret the data of an individual study by unspecified expert judgement than by quantitative statistical methods. Thus, there must be merit in the use of procedures that are explicit, repeatable by others, and deemed rational by many, even when the circumstances of each single study are rife with potential distortions. If so, it ought to be also peculiar to argue that meta-analysis in road safety should rely solely on expert judgement. The quantitative method presented in this section can help to infuse into the meta-analytic process a similar kind of explicit discipline as the usual statistical methods provide for the interpretation of a single study.

Selected Books on Meta-Analysis. Glass, G.V., McGraw, B. and Smith, M.L. , (1981). *Meta-analysis in social research.* Sage, Beverly Hills, California. Light, R.J. and Pillemer, D.B., (1984). *Summing up: The science of reviewing research.* Harvard University Press, Cambridge, Massachusetts. Rosenthal, R., (1984). *Meta-analytic procedures for social research.* Sage, Beverly Hills, California. Hedges, L.V. and Olkin, I., (1985). *Statistical methods for meta-analysis.* Academic Press, Orlando, Florida. Cooper, H.M., (1989). *Integrating research: A guide for literature reviews.* Sage, Beverly Hills, California. Wachter, K. W., and Straf, M. L. (eds.), (1992). *The future of meta-analysis.* Russel Sage Foundation, New York. Hunter, J.E. and Schmidt, F.L., (1990). *Methods of meta-analysis.* Sage Publications. London. Petitti, D.B., (1993). *Meta-analysis, decision analysis, and cost-effectiveness analysis.* Methods for quantitative synthesis in medicine. Monographs in epidemiology and biostatistics, Volume 24. Oxford University Press, New York, Oxford.

10.4 CHAPTER SUMMARY

This chapter revolves around the recognition that the same treatment when applied in different circumstances will affect safety to a different extent. We need to know how large is this variability of safety effect, what are the circumstances in which the treatment is most beneficial and in which conditions its benefit is smallest.

The need to estimate not only the mean safety effect but also its variance led to the expanded 'four-step' in Section 10.1. We now imagine a large population of entities that could be treated. The treated entities for which we have data are thought of as a representative sample of this population. What we estimate to be the effect of the treatment on the entities of this sample, is then used to make inferences about what the safety effect would be on entities from the imagined population, if they we similarly treated. The familiar estimate of $\bar{\theta}$ is now used to estimate the mean of the θ's of the imagined population. The estimate $V\hat{A}R\{\theta\}$ in Equation 10.1 is used to estimate the variance of the θ's in the imagined population. Equation 10.1 is the main result of this chapter.

It gives a simple way to estimate the VAR{θ} when we have estimates $\hat{\theta}(1)$, $\hat{\theta}(2)$, . . . , $\hat{\theta}(n)$ and $V\hat{A}R\{\hat{\theta}(1)\}$, $V\hat{A}R\{\hat{\theta}(2)\}$, . . . , $V\hat{A}R\{\hat{\theta}(n)\}$ for the 'n' entities of the sample.

I use data about the installation of raised pavement markers on 86 road sections in Texas to illustrate the procedure. The assumptions are that raised pavement makers affect only nighttime accidents, and that daytime accidents on the same sections can serve as an ideal comparison group. Data analysis leads to the conclusion that in the imagined population of future applications of raised pavement markers in similar circumstances one should expect an average increase of nighttime accidents by 17.6%. The standard deviation of θ's in this imagined population is 32.8%. Thus, while on the whole this treatment reduces nighttime safety, in about one third of the cases (see figure 10.1) safety is enhanced. The next step ought to be an examination of where the treatment is most beneficial and where most harmful.

The ability to estimate VAR{θ} proves useful also when we have a series of estimates $\hat{\theta}(j)$ and of the variance of these estimates $V\hat{A}R\{\hat{\theta}(j)\}$ from j=1, 2, . . . , n studies in which the same treatment has been implemented. The quantitative means for extracting the lessons of several studies have been given the name meta-analysis. I use data from studies of the safety effect of illumination on pedestrian accidents in urban areas for illustration. Analysis leads to the conclusion that in the imagined population of illuminations in urban areas pedestrian accidents at night are reduced, on the average, by 54%. The $\sigma\{\theta\}$ in this imagined population is 12%.

Chapters 6 to 9 were about the common ways to obtain the estimates $\hat{\theta}\{j\}$ and $V\hat{A}R\{\hat{\theta}(j)\}$of a treatment. In this chapter I show that, no matter how these estimates are obtained, they can be used to estimate the mean and the variance of the θ's in future applications of the treatment. However, all the methods considered so far are attempts to adapt some common experimental designs to the realities of observational studies. These adaptations were not entirely satisfactory. The next part of the manuscript is about new approaches to the interpretation of observational Before-After studies.

PART III

ELEMENTS OF A NEW APPROACH

In the first part of this monograph I have discussed such general issues as the logic of Before-After studies, the definition of safety, and the central role that prediction plays in the estimation of the safety effect of treatments. In the second part I dwelled on some conventional approaches that reflect the heritage of the physics laboratory (the Naive method) or the ancestry of randomized statistical experiments (the Comparison-Group method). These conventional methods do not fit well the usual circumstances of observational studies about the effect of some treatment on road safety. Neither is it possible to keep all conditions constant, as in the laboratory, nor do we often have the luxury of a random assignment of entities to sufficiently large treatment and control groups, as is necessary for statistical experimentation. Accordingly, the tools for best interpreting an observational study are not those suited for the analysis of a statistical experiment.

The tension between the reality of observational studies and the conventional approaches to their interpretation is manifest in three issues:

First, in all the approaches discussed in Part II the assumption was, that the count of accidents in the 'before' period (K) is a sensible estimate of the expected value E{K}. However, in an observational study one may not assume that the past accident history has not been considered when the decision was made to treat or not to treat an entity. Often the opposite is true. Sensible people apply treatments when they seem necessary. This leads to the acute problem known as the 'selection bias' or the 'regression-to-mean bias'. If the count of 'before' accidents (K) influenced the decision to 'treat or not to treat', K is a biased estimate of E{K}. The neglect of this bias in many published studies led to exaggerated estimates of effect for many interventions and is still polluting the professional lore. A method has to be devised to purge the results of observational studies from the threat of the regression-to-mean bias.

Second, the current habit is to use short 'before' periods. The received wisdom is for the 'before' period not to exceed three years. Perhaps this makes the naive assumption that the 'before' accident count is a good prediction of 'what would have happened after' seem less implausible. But this cutoff is arbitrary. The 'before' period has a natural end, but not a natural beginning. To discard accident counts that are more than three years old amounts to a loss of potentially useful information. There is another and more serious problem associated with the use of a short 'before' period; in a few years of accident counts one cannot discern the existence of a time trend. To assume that there is no time trend just because it cannot be seen in a few years of accident counts is wrong. Surely one can predict better if the accident counts for more than three years is used and if the possible existence of a time trend is not disregarded.

Third, the approaches of Part II use impoverished methods to predict what safety would have been in the 'after' period had there been no intervention. Changes in traffic flow are usually accounted for assuming proportionality, even though the relationship between traffic flow and safety is seldom linear. Changes in other factors are accounted for by using a comparison group, even though the comparison group is neither the only nor the best device for prediction. Better tools for predicting can be devised.

These issues demand attention. Accordingly, the third part of this monograph is devoted to approaches that fit the realities of observational Before-After studies better. I will deal with the three issues seriatim. In Chapter 11 it will be shown that the accident record of an entity in the before period is not a good estimate of its safety. I will argue that the Empirical Bayes (EB) approach to estimation not only solves the regression-to-mean problem but also yields more precise estimates. In Chapter 12 estimation will be freed from the constraint of a fixed-duration before period. The key role played by multivariable models which express the safety of an entity as a function of its observable traits will be examined. I will then return to the central issue - that of estimating what the effect of some intervention is. Finally, I will provide an extended case study illustrating how this approach has been used to estimate the effect of resurfacing rural roads.

CHAPTER 11

BACK TO THE STARTING POINT
THE EMPIRICAL BAYES APPROACH

To estimate how some treatment affected safety, we need to **predict** what the expected number of accidents would have been in the 'after' period had the treatment not been implemented, and to compare this prediction with what safety in the 'after' period was, with the treatment in place[1]. As noted in Section 5.1, there are many ways to predict. However, all methods of prediction can be thought to consist of two consecutive steps.

a. First one forms the foundation for the prediction by estimating what the expected frequency of target accidents in the 'before' period **was**.

b. Then, using the foundation formed in step 'a', one predicts how the expected number of target accidents **would have changed** from the 'before' to the 'after' period as a result of changes in traffic, weather and all other factors, save the treatment.

Step 'a' is firmly in the domain of statistics. Step 'b' is inductive in nature; it amounts to modeling, extrapolation or prophecy based on a belief in the continuation of past regularities.

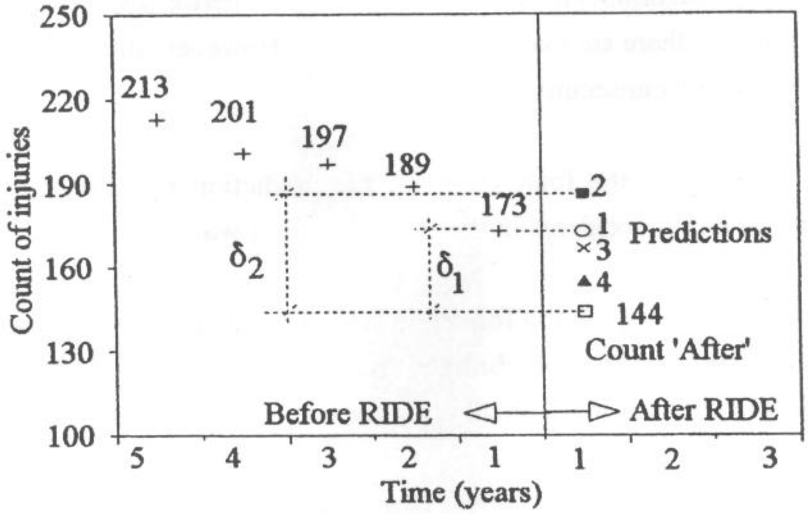
11.1 THE SHAKY FOUNDATION AND HOW TO SHORE IT UP

In Chapters 7, 8 and 9 the tacit assumption was that step 'a' is not problematic. In both the 'Naive' and the 'Comparison Group' approaches, the key assumption of step 'a' was[2] that for any treated entity, the count of 'before' accidents, K, is a sensible estimate of its expected accident count κ. There are two problems with this assumption. First, the assumption is true only if the accident history of the entity had nothing to do with the reasons for it being treated. In observational studies,

[1] See Chapter 5.

[2] See the estimate of π in Table 7.1, Equation 7.1 and Table 9.3.

this may be untrue. Intersections may be equipped with traffic signals if the accident warrant[1] is met; crossbucks at grade-crossings are often changed to flashers or gates in the aftermath of a severe car-train collision; shoulder-parking may be prohibited where there were many accidents with parked vehicles; drivers are called in for training if they had many convictions; roads may be resurfaced if a wet-weather accident problem is noted; speed enforcement is increased where speeding is cited as cause of accidents; etc. Thus, in an observational study there is likely to be a link between the decision to treat an entity and its accident history. This link causes a so called 'selection bias' or 'regression-to-mean' (RTM) bias. It makes K a biased estimate of κ as will be amply demonstrated in Section 11.2. The second problem with the assumption that K is a good estimate of κ becomes evident after a glance at Figure 5.1 reproduced here.



Accident counts, predictions and estimates of change.
(Same as Figure 5.1.)

The 'before' period ends when treatment is applied. However, there is no fixed time when the 'before' period begins. Unlike statistical experiments, observational Before-After studies do not have a natural 'before' period of fixed duration. But, if we do not speak of some fixed duration, then what is κ? Is it the expected accident frequency one year before treatment? Is it the average for the last three years? Since every entity exists in the real world of changing traffic, vehicles, habits, weather; etc., the κ pertaining to the first year before the treatment is bound to differ from the κ pertaining to the previous year, and both differ from the preceding κ's. The real task (for step 'a') is not to estimate 'the' κ for some arbitrarily selected 'before' period as has been done in the Naive

[1] Thus, e.g., in the Manual on Uniform Traffic Control Devices (FHWA, 1988) a traffic signal is deemed warranted if five or more correctable accidents occurred at an intersection in 12 months.

or C-G studies of Part II. The task is to estimate a sequence κ_1 , κ_2 , κ_3 , . . . for as many years into the past as is feasible and helpful. It is this sequence of κ's that is to be the foundation for the projection that is fashioned later, in step 'b'. To remedy both problems, the use of the Empirical Bayes (EB) approach to estimation will be advocated. Its merits are three. First, the EB approach to estimation helps to deal with the regression-to-mean (RTM) bias. Second, EB estimates tend to be more precise than estimates used earlier in Part II. Third, the EB approach allows the estimation of the entire time series κ_1 , κ_2 , κ_3, . . . as required. The essence of the EB approach is that two separate pieces of information are used to estimate the safety of a certain entity,

a. the accident history of that entity and,
b. what is known about the safety of other entities with similar traits.

While such an approach to estimation makes common sense, it is not what is normally done. Thus, e.g., in all the conventional approaches which make up Part II, only the accident history of an entity served as data for estimating its safety; knowledge of the safety of entities with similar traits (such as traffic, geometry, control devices etc.) was not utilized[1]. Textbooks and courses on probability tend to reinforce the peculiar notion that expected frequencies are to be estimated solely from the outcomes of repeated trials. The fact that one learns more about the expected frequency of heads for a coin by looking at it, than by tossing it a hundred times is seldom discussed[2]. Similarly, in estimating the safety of a certain grade crossing, one should consider not only its own history of accidents (which is bound to be sparse) but also what is the accident experience at other grade crossings with similar traffic and warning devices.

Because what I advocate departs from what is usually done, it is essential to begin by showing how, why, and when the conventional approaches to estimation lead to the RTM (regression-to-mean) bias (Section 11.2). In Section 11.3 I discuss the logic of the EB method in non-mathematical terms and point out what difficulties it brings about. Its statistical justification and mathematical aspects are introduced in Section 11.4. In Section 11.5 I show how to obtain the requisite 'weights' to make the EB method operational. Since the proof of this pudding is in applications, in Section 11.6 I show how the EB approach cures the regression-to-mean problem and how it is of use in answering several questions that are of practical interest. In this chapter I will still proceed as if it made sense to have a fixed-duration 'before' period. This will render the results useful to those who wish to use a fixed-duration 'before' period in their work. The same results will help later (in Chapter 12) to deal with the second problem, that of freeing the method of prediction from the artificial constraint of a fixed-duration 'before' period.

---

[1] Except for making corrections due to change in some causal factors.

[2] A look at the coin will show that it is ordinary, and brings into estimation the knowledge that ordinary coins tend to be fair.

11.2 THE REGRESSION-TO-MEAN PHENOMENON

Early in Chapter 7, selection bias and regression-to-mean bias were mentioned as one of the four reasons for doubting the results of a Naive Before-After study. A real example is in Section 7.4, where a red flag was raised about the claim of reduction in right-angle accidents that has been attributed to some signal changes because the author said that "large safety benefits" were found at intersections "experiencing a large number of right-angle collisions prior to the" intervention. This is precisely the circumstance in which RTM arises.

The phrases 'selection bias' and 'regression to the mean bias' evoke images of a slightly mystical phenomenon, one that is difficult to understand by common sense. In fact it can be explained quite simply. If an entity is treated because its 'before' accident count (K) was abnormally high or unusually low, then the same K cannot possibly be a good estimate of κ. After all, κ is the 'expected' value of K; that is, the very embodiment of a usual, normal, or average accident count. One cannot estimate what is normal and usual by using accidents counts that are abnormal or unusual. To do so would lead to an obvious bias. If the entity has been selected because it had an unusually high number of accidents, K would tend to overestimate κ.

The RTM bias may be present even if entities were not selected **because** of abnormally high or unusually low accident counts. Even if entities with unusually high or low accident counts just **happened to be selected for treatment** in a specific case, even then K is not a good estimate of κ - in this case. This is why, even when comparison groups were used, as in Section 9.4, and when data from all three cities seemed to point in the same direction, I could not say by how much safety was diminished due to the conversion from STOP to YIELD control. The source of my reluctance was that the treated intersections seemed to have more accidents than other intersections of the same kind. Whether by design or by chance, sites with unusually large accident counts seem to have been converted from STOP to YIELD control. Since this has not been recognized and accounted for in the analysis, the (detrimental) effect of the STOP to YIELD conversion is likely to be even larger than what it has been estimated to be. This is why the basic premise of the statistical analysis of Part II, namely, that K is a good estimate of κ, was said to be "in danger of being wrong."

At this point I have to explain in more detail when K is not a good estimate of κ and why. The most common circumstance is when entities are selected for improvement because of their bad accident record as shown in the following example. The example is hypothetical so as to illustrate the phenomenon with clarity. Examples based on empirical evidence will be furnished later.

Numerical Example 11.1. In the aftermath of a warrant.

Imagine a city with 100 nearly identical intersections equipped with stop signs. Assume that the expected frequency of correctable accidents at each intersection is κ=3 accidents/year. Even though all have the same κ, in a specific year, some intersections will have 0 accidents, some will have 1, etc. If the count K of accidents is Poisson distributed, one can compute how many of these 100 intersections are expected to record 0, 1, 2, . . . accidents in a year. This computation is tabulated below:

Table 11.1. Expected number of intersections with K accidents in a certain year.

	Count of accidents (K)	Probability that an intersection has K accidents	Expected number of intersections with K accidents
MUTCD warrant not met	0	.0498	5
	1	.1494	15
	2	.2240	22
	3	.2240	22
	4	.1680	17
MUTCD warrant met	5	.1008	10
	6	.0504	5
	7	.0216	2
	8	.0081	1
	≥9	.0040	1

Thus, even though κ=3 for all 100 intersections, in any 12-month period one should expect 10 intersections to record 5 correctable accidents, 5 intersections to record 6 such accidents, etc. The Manual on Uniform Traffic Control Devices (FHWA, 1988, 4C-6) suggests that signalization may be considered if an intersection recorded 5 or more correctable accidents in a 12 months period. In an average year, of the 100 intersections that are all equally safe, 19 would meet the MUTCD warrant (see the bottom part of Table 11.1).

Suppose that at the end of such a year, the stop signs at these 19 intersections were replaced by traffic signals and that signalization reduced the expected number of correctable accidents by 10%. That is, after signalization, κ=2.7 correctable accidents per year are expected to occur at each of the 19 signalized intersections.

Numerical Example 11.1 (continued).

The conscientious traffic engineer decides to conduct a Naive study with one-year 'before' and 'after' periods. To keep things simple, assume that traffic flow and all other factors were identical in the 'before' and 'after' years. In such conditions, there is ostensibly no reason to doubt the results of a Naive study. Should the Naive study be expected to find the truth, namely a 10% improvement in safety?

The answer is: "No". For the 19 newly signalized intersections the number of 'before' accidents was $111(=5\times10+6\times5+...+1\times8+1\times9)$. In the 'after' year each newly signalized intersection is expected to record 2.7 correctable accidents. Thus, for the 19 intersections we expect in the 'after' year to count $19\times2.7\approx51$ correctable accidents. Now, by Equation 7.3, $\hat{\theta}=(51/111)/(1+1/111)=0.46$. This appears to be a 54% reduction. However, we know that the reduction was only of 10%! Thus, RTM makes the effect of signalization appear much larger than it actually is. This is due entirely to the fact that, as advised by the MUTCD warrant, intersections with 5 or more correctable accidents were selected to be signalized.

In sum, although in reality the change in safety was only of 10%, a correctly performed Naive Before-After study is expected to show a 54% change. The same bias would beset the results of a C-G study. Thus, the bias cannot be avoided by using a regular comparison group[1]. The reason for the failure of the Naive Before-After study to estimate correctly the effect of signalization on safety in this case is now evident:

> The count of 'before' accidents (K=111) is not a good estimate of what κ in the 'before' period was.

By the assumption of this example, each intersection was expected to record 3 correctable accidents before signalization. Thus, the long-term average over many years for these 19

[1] To begin with, a comparison group is not needed here because I have postulated that "traffic flow and all other factors were identical in the 'before' and 'after' years." However, were a comparison group used, it would, on the average, show 3 correctable accidents for the 'before' and also for the 'after' period. Thus, the RTM bias would remain intact. The only comparison group that could account for the RTM bias, would be the one created by a random assignment to 'treatment' and 'control' of the 19 selected intersection. Were this possible, we would have a statistical experiment, not an observational study. Thus, in an observational study, the use of a comparison group is not a device for removing the RTM bias.

intersections would be 3×19=57 accidents. But, in the specific year considered here, these 19 intersections had an unusually high number of accidents (111). Indeed, this is why they were identified by the Manual on Uniform Traffic Control Devices (FHWA, 1988) as warranting signalization. Obviously, under these circumstances, it is incorrect to assume that K is a sensible estimate of κ.

A hypothetical example may not be as convincing as real data. Therefore, I provide Table 11.2 which is based on the count of accidents occurring during the years 1974 and 1975 at 1142 intersections in San Francisco. All intersections in this group had stop signs on the two approaches carrying the lesser flows. Column 1 gives the number of intersections [n(K)] for which the count of accidents [K] in 1974 was 0, 1, 2 . . . as shown in column 2. Column 3 gives the average of the count of accidents per intersection [avg(K)] for the same n(K) intersections during 1975.

Table 11.2. Accident count at 1142 intersections - 1974/1975.

1 Number of Intersections n(K)	2 Number of Accidents per intersection in 1974 K	3 Average Number of Accidents per intersection in 1975 avg(K)
553	0	0.54
296	1	0.97
144	2	1.53
65	3	1.97
31	4	2.10
21	5	3.24
9	6	5.67
13	7	4.69
5	8	3.80
2	9*	6.50

* In addition, 2 intersections had 13 accidents, one had 16.

Suppose that we used K to estimate κ, as is done in the Naive and in the C-G methods. This would mean that if an intersection registered, say, K=3 accidents in 1974, and had it remained largely unchanged for 1975, we would assert that 3 is a correct and sensible estimate of the accident count for 1975. This assertion applies to each of the 65 intersections in the group with K=3.

Therefore, while for any intersection in this group the 1975 count of accidents may be different from 3, the average count for the group of 65 intersections should be quite close[1] to 3.

However, inspection of Table 11.2 reveals that intersections which each registered 3 accidents in 1974, had 1.97 accidents on the average in 1975! Similar discrepancies between the entries of columns 2 and 3 exist for all values of K (except for K=1 which will turn out to be the rule-confirming exception). These discrepancies cannot be reasonably attributed to chance. Nor are they likely to reflect a sudden, large and peculiarly systematic change between these two years[2]. I observe therefore that, in this case, the count of accidents in 1974 is not a good indication of the average count in 1975 for any value of K (except for K=1). Empirical evidence flies in the face of what is thought to be naturally correct. If K is not a good estimate of the mean number of accidents for a group of intersections, it cannot be a good estimate for any intersection in that group.

Is it possible, that what we see in Table 11.2 is because K is for one year only, that was a longer accident record used, K would prove to be a good estimate of κ? The answer is "No." To demonstrate this, I list in column 1 of Table 11.3 the number of San Francisco intersections which recorded a total of K=0, 1, 2, . . . ,9 accidents during the three years 1974-1976.

Table 11.3. Accident counts at 1072 intersections with up to 9 accidents in 1974-77.

1 Number of Intersections n(K)	2 Accidents per Intersection in 1974-1976 K	3 Average Accidents per Intersection in 1977 avg(K)
256	0	0.25
218	1	0.55
173	2	0.70
121	3	1.04
97	4	1.08
70	5	1.33
54	6	1.56
32	7	2.25
29	8	1.62
22	9	2.50

[1] If these intersections had κ=3 accidents/year and the accident count was Poisson distributed, the standard deviation of avg(K) would be $\sqrt{3/65}$=0.21.

[2] The total number of accidents at these intersections was 1253 in 1974 and 1216 in 1975.

In column 3 I show the average number of accidents recorded at these intersections in 1977. If K was a good estimate of κ, one should expect a similarity between the entries of column 3 and 1/3 of column 2. The discrepancies, however, are as regular and persistent as in Table 11.2.

A visually striking illustration of the regression-to-mean at work is shown in Figure 11.1. In it I trace the average number of accidents in the years 1975-77 for intersections which in 1974 had 0, 1, 2, . . . accidents. I noted that there is a difference between K and avg(K) in columns 2 and 3 of Tables 11.2 and 11.3. A similar difference is evident in Figure 11.1 between the K in 1974 and the avg(K) in 1975, 1976 and 1977. Note that there is a distinct pattern to this difference. This data shows that in 1974, 1253 accidents occurred at 1142 intersections. Thus, the population mean $\bar{K}$=1253/1142=1.1 accidents/year. When K<1.1, then K<avg(K)<1.1; also when K>1.1 then 1.1<avg(K)<K. What we observe here is the phenomenon of regression-to-the-mean.

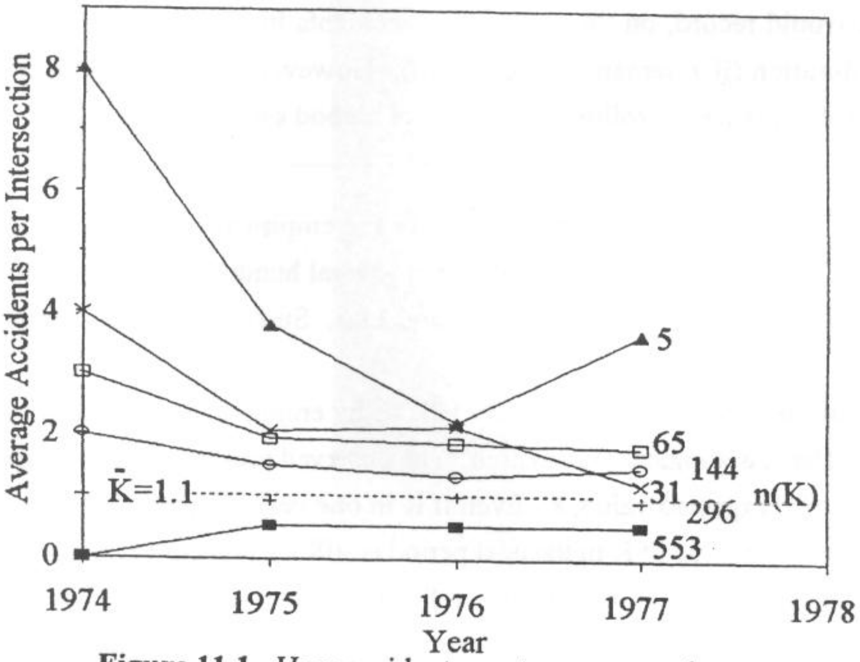


Figure 11.1. How accident counts regress to the mean.

Sir Francis Galton, the noted meteorologist, biologist and statistician, reported in 1877 that the offspring of tall parents are, on the average, shorter than their progenitors (and vice versa). Kendall and Buckland (1967) attribute to Galton the term 'regression' in its meaning of a tendency to 'return toward' the mean. In Galton's case as in ours, what happened 'before' (the stature of the parent or the count of accidents 'before') turned out to be an imperfect and biased indication of what happened 'after' (the height of the offspring or the count of accidents 'after'). The term 'regression' has in our days acquired a new meaning. However, the original phenomenon noticed by Galton more than a century ago has been consistently observed since his time in a variety of situations. Thus, e.g., Efron and Morris (1977) use an example of baseball batting averages, Hauer and Persaud

(1983) show its occurrence in golf scores, traffic law violations and, of course, counts of road accidents. In all cases one finds that what happened 'after' is, on the average, not what happened 'before'.

In our earlier work (Hauer and Persaud, 1983) we examined a large number of similar data sets pertaining to drivers, road sections and intersections in various countries. Just as in Tables 11.2 and 11.3, we found in all cases large and systematic discrepancies between K (in its capacity as the estimator of κ) and between avg(K) (in its role as the estimate of the mean of the average κ for the group of entities which have registered the same K). It follows, that the results in Tables 11.2 and 11.3 are not an exception or aberration. They serve here as representatives of a phenomenon which is general. I can distill the claim into the following chain of reasoning:

> If K was a good estimator of κ then, entities which recorded K accidents in one period would record, on the average, K accidents in a subsequent period of equal duration (if κ remained unchanged). However, this is not born out by empirical evidence. It follows that K is not a good estimator of κ.

The reader is invited to independently verify the empirical basis of the claim. All that is needed are at least two years of accident data about several hundred 'entities' (intersections, road sections, drivers etc.) arranged as in Tables 11.2 and 11.3. Such data sets are ubiquitous.

What nature so clearly and consistently tells us by empirical facts has not only a name and historical roots, it also has a logical explanation. The observed accident count, K, always fluctuates around some unknown expected value, κ. Even if K in one period happens to be larger than κ, the best guess about the magnitude of K in the next period is still κ. This comes close to the essence and meaning of 'expected value'. Conversely, if the first observed value of K happens to be a 'down-fluctuation' or an 'up-fluctuation', a return to the mean should be expected for the next observation. This is how regression-to-the-mean works at the level of a single entity. When single entities are aggregated into groups which have K in common as has been done in Tables 11.2 and 11.3, there is a good chance than an entity included in a group with a large K happened to be observed in an 'up-fluctuation'. Therefore it is expected to return to its mean during the next observation period. This is why we find that for the groups with large K, avg(K)<K and vice versa.

If K is not a good estimate of κ, a better method of estimation needs to be suggested. The Empirical Bayes approach introduced next will fulfil this need.

11.3 TWO CLUES TO SAFETY

In the preceding section I have demonstrated that the count of accidents for an entity (K) is a biased estimate of its expected number of accidents (κ) when the K has something to do with the reason why the entity is being considered. In such circumstances a different approach to estimation is needed. The purpose of this section is to introduce the Empirical Bayes (EB) approach as a matter of good sense. Mathematical and statistical issues will be discussed in subsequent sections.

The 'safety' of an entity has been defined to mean the number of accidents by kind and severity, expected to involve that entity (or to occur on it), per unit of time, in a certain period. The word 'expected' refers to a long-term average which would materialize if it were possible to keep constant traffic flow, driver demography, vehicle characteristics and all other relevant conditions of the period, as well as all the traits and properties of the entity. Thus, e.g., we might be interested in the safety in 1984 of Mr. Smith, who in that year was 22 years old and who had an accident in 1982; or we might wish to estimate the safety of a specific rail-highway grade crossing in 1980, which served an average of 800 vehicles/day, two trains/day, had one track, was equipped with crossbucks and recorded two accidents between 1970 and 1980.

There are two kinds of clues to the safety of an entity. **Clues of the first kind are contained in traits** such as gender, age, traffic or geometry. Statistical evidence suggests, e.g., that young males are less safe than middle-age females. For this reason we might expect Mr. Smith (age 22) to be involved in more accidents per unit of time than Mrs. Jones (age 50). In this case, the traits 'gender' and 'age' when coupled with accident statistics about groups defined by these traits, serve as clues to safety. Similarly, we expect more accidents per unit of time on a kilometer of an urban arterial road than on a kilometer of a rural two-lane road. The traits here are 'rural', 'urban', 'arterial' and 'two-lane'.

Clues of the second kind to the safety of an entity are derived from the **history of accident occurrence** for the entity of interest. If Mr. Smith had one accident in the past four years while Mrs. Jones had three, this too is a useful clue to their safety. Similarly, the accident history of a road section or intersection contains information about their safety.

It is not yet obvious how the two kinds of clues can be combined into one quantitative estimate of safety. How this can and should be done will be discussed shortly. Nevertheless, it is clear that to disregard one or the other clue is contrary to good sense. Yet this is precisely what we do as a matter of routine. In all the conventional methods of Part II, safety has always been estimated using only clues of the second kind - the accident counts.

The essence of the EB (Empirical Bayes) approach is to use both clues for safety estimation. Clues of the first kind are those that are based on the traits of the entity and on the safety of the group of entities sharing these traits. This is the 'mean safety' toward which the individual estimates were seen to 'regress'. Clues of the second kind, those derived from the accident counts of an entity, determine by how much the κ of a specific entity deviates from what is mean for the entire group.

To prepare the ground for the procedure by which both clues will be combined, I have to coin a new concept and to explain its meaning. Clues of the first kind enable one to make informative statements about the safety of a specific entity only if some knowledge exists about the safety of a group of entities with similar traits. The group of entities with 'similar traits' will be called the **'reference population'**. To illustrate, consider the safety of Smith in 1984 when the only thing known about this driver is that she or he was licensed to drive in Ontario, and that in Ontario there were 0.0606 accidents per licensed driver in that year, on the average. When I say that I estimate Smith's safety in 1984 to be 0.0606 accidents/year, the following long-winded reasoning applies:

> The population of persons licensed to drive in Ontario in 1984 is the **reference population** for Smith because nothing is known to distinguish Smith from others in this population. Since they are indistinguishable, the estimate of safety for any member of this population, including Smith, should be the same. If estimating with the least squared error is desirable, it is best to use the mean. Here, the mean is thought to be 0.0606 accidents/year. This is why it is proper to assert that 0.0606 is the estimate of safety for Smith.

If it were known, in addition, that Smith was 22, this added trait would define a new reference population. It is a subset of the earlier reference population and contains only those Ontario drivers who in 1984 were 22 years old. For this new reference population official statistics show the average number of accidents/year to be 0.0840. Following the same reasoning as before, this is the estimate for a 22-year-old Smith. If it were also known that Smith is male, yet a smaller reference population is defined, members of which share the traits of being a driver in Ontario in 1984, being 22, and being male. For this population there was an average of 0.1152 accidents per driver per year. If nothing else is known about Mr. Smith, 0.1152 must be the estimate of his safety in 1984. One should not find it disconcerting that the estimate of Smith's safety changes as more of his traits are revealed. This is merely a manifestation of the property of all estimators, namely, that they depend on the data we have. In sum,

a reference population of entities is the group of entities that share the same set of traits as the entity in the safety of which we have an interest.

It is this set of traits that determines which entities belong to the reference population. The concept of 'reference population' introduced in this section will play a pivotal role in the EB method. What is said about the safety of Mr. Smith depends on what is known about his traits, for it is these traits that define the reference population. To say something about the safety of Smith we have to know what was the safety of entities in the reference population. This brings about a string of difficulties which are described and discussed in the endnote to this chapter.

In a driver's life accidents are relatively rare. Nine in ten drivers in Smith's reference population would have no accident in a given year. It therefore makes common sense that the average of a reference population be an influential clue to Smith's safety. At the other extreme, if we are interested in the safety of a city or a country, where many thousands of accidents occur every year, the main clue to their safety are the actual accident counts, not the average of some reference population of similar cities or countries. Between these two extremes is the domain of applications: intersections, road sections, grade crossings, etc.; entities on which accidents are not so frequent that the clues deriving from traits can be disregarded, nor are accidents so rare that they can be disregarded and the mean of a reference population used to estimate safety.

What we know so far is that Smith is male, 22 years old, licensed to drive in Ontario and that such people had, on the average, 0.1152 accidents in 1984. But we also know that Mr. Smith had one accident in 1982. The next question is how this knowledge of Smith's accident record should affect what we estimate his safety in 1984 to be? This question is tackled in the next section.

11.4 THE MATHEMATICS OF MIXING THE TWO CLUES

Accepting the notion that all available evidence should be brought to bear on the task of estimating the safety of an entity is easy. It is also plain to see that there is useful information in both kinds of clues:

1. Clues that reside in the traits of the entity, for these traits make it similar to other entities about the safety of which we have data;
2. Clues that reside in the accident record of the entity, for it is a reflection of its safety.

However, it is far from self-evident how to join these two diverse pieces of information into one estimate. The art of estimating safety using both kinds of clues is described in a handful of articles (Abbess et al, 1981; Hauer and Persaud, 1983; Hauer, 1986a,b; Morris et al., 1989; Hauer, 1992b). The main results are given below.

For a certain entity we wish to estimate κ, its expected number of accidents of some kind and pertaining to a given period. What we know about this entity are two things: First, that by its traits the entity belongs to a certain reference population - the entities which have κ's with a mean $E\{\kappa\}$ and variance $VAR\{\kappa\}$; Second, that in the period of interest the entity recorded K accidents. These two pieces of information need to be joined. Their joining into one estimate of the κ of 'our' entity is based on the following argument.

Consider the entities of the reference population. Some entities recorded no accidents, some recorded one, some two and so on. Focus now on those entities of the reference population that have all recorded exactly K accidents in a period of interest. Let $E\{\kappa|K\}$ and $VAR\{\kappa|K\}$ denote the mean and the variance of the κ's in this sub-population[1]. I now resort to the same reasoning as in the preceding section. Since 'our' entity has the same traits and the same accident record as all the other entities of this sub-population, its κ can be, with equal probability, any of the κ's in this subpopulation of entities that all recorded K accidents. Therefore, **the best estimate of the κ of 'our' entity is $E\{\kappa|K\}$, and the variance of this estimate is $VAR\{\kappa|K\}$**. Thus, we are looking for the logic and the statistical tools by which, from information about $E\{\kappa\}$, $VAR\{\kappa\}$ and K, one can forge the items of interest, namely, $E\{\kappa|K\}$ and $VAR\{\kappa|K\}$. This line of reasoning is depicted in Figure 11.2.

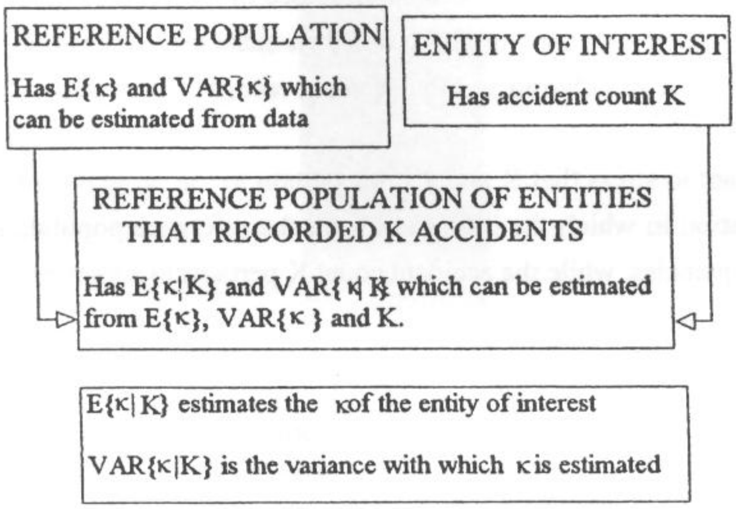


Figure 11.2. The logic of estimation.

[1] The vertical bar signifies the 'given that' condition and the text behind it says what is known or 'given'. Thus, e.g., $E\{\kappa|K\}$ is the expected value of κ when it is known (or given) that the entity recorded K accidents.

Naturally, E{κ|K} will be some mix of the two constituent elements: the E{κ} of the reference population and the K of the entity at hand, assuming some value between the two. Thus,

$$E\{\kappa|K\} = \alpha E\{\kappa\} + (1-\alpha)K \qquad \ldots 11.1$$

In this, α is a number between 0 and 1 that needs to be chosen. If α is chosen to be near 1, then the κ of the entity of interest (as estimated by E{κ|K}) is close to the mean of its reference population, E{κ}; if α is chosen to be near 0 then the κ of the entity of interest will reflect mainly the recorded count of accidents, K. The question is how to choose the 'weight' α.

In the derivations (at the end of the section) I show that if one wishes to estimate the κ of the entity with maximum precision, then

$$\alpha = \frac{1}{1 + \dfrac{VAR\{\kappa\}}{E\{\kappa\}}} \qquad \ldots 11.2a$$

Note that α is only a function of the mean and the variance of the κ's in the reference population and is always a number between 0 and 1, as required. Exactly how the E{κ} and VAR{κ} of the reference population are to be estimated will be discussed in the next section. Thus, to estimate the κ of a certain entity, first obtain the 'weight' α and then make use of Equation 11.1 which combines the mean of the reference population and the accident record of the entity.

It is important to stress that K and κ must pertain to the same period. In practice one may encounter the situation in which the information for the reference population is in terms of, say, annual accident frequencies, while the accident count K pertains to a two-three year 'before' period. If so,

$$\alpha = \frac{1}{1 + r\dfrac{VAR\{\kappa\}}{E\{\kappa\}}} \qquad \ldots 11.2b$$

in which the ration r is the number of years to which K pertains divided by the number of years to which κ pertains. This accomplishes one part of what we set out to do. Information about E{κ}, VAR{κ} and K, can now be used to estimate E{κ|K} which, in turn, serves to estimate the κ of 'our' entity.

Numerical Example 11.2. Safety of a grade crossing.

Consider a highway-rail grade crossing which is in an urban area, has a single track, is used by 2 trains per day, 550 cars per day and is signed by cross-bucks. In the five years, 1981 through 1985, it has recorded two accidents. What is the estimate of its κ during that period?
Answer: It will be shown later (in Numerical Example 11.6) that in a reference population of such crossings, $\hat{E}\{\kappa\}=0.0239$ accidents/year and $V\hat{A}R\{\kappa\}=0.0011$ [accidents/year]^2. For the five years, $\hat{E}\{\kappa\}=5\times0.0239=0.1195$ accidents and $V\hat{A}R\{\kappa\}=5^2\times0.0011=0.0275$ [accid.]^2. Therefore, by Equation 11.2a, $\hat{\alpha}=1/(1+0.0275/0.1195)=0.81$. Equivalently, by Equation 11.2b, $\hat{\alpha}=1/(1+5\times0.0011/0.0239)=0.81$. Now, using Equation 11.1, $\hat{\kappa}=0.81\times0.1195+0.19\times2= 0.10+0.38=0.48$ accidents in five years.

The essence of the EB method is now in plain view. The two clues to κ (0.1195 describing an average in the reference population, and 2 being the accident count on the entity of interest) are combined. The weight attached to each clue depends only on the ratio $VAR\{\kappa\}/E\{\kappa\}$. Note that if all κ's in the reference population were the same, then, in Equation 11.2, $VAR\{\kappa\}=0$. If so, $\alpha=1$. In this case, the count of accidents does not influence the estimate. While at first sight this may seem odd, upon reflection it is entirely appropriate. If all entities with these traits are known to have the same κ, which $VAR\{\kappa\}=0$ implies, then, whatever the accident count happens to be, $E\{\kappa\}$ must be the estimate of the κ for the entity under scrutiny. Conversely, if the κ's in the reference population are so diverse that $VAR\{\kappa\}>>E\{\kappa\}$, α would be very small. If so, what is known about the reference population exerts little influence on the estimate. This is also as should be. Thus, while the estimator specified by Equations 11.1 and 11.2 has been deduced by logic from specified premises, it is seen to mix the two kinds of clues in a manner that makes good sense.

Yet another aspect of this estimator deserves noting. When discussing the regression-to-mean phenomenon in Section 11.2, I pointed to the distinctive pattern in which any K of 1974 shrinks toward the mean in subsequent years (see Figure 11.1). In general, if K in one year was larger than the mean for the population (in Figure 11.1 the sample mean for 1974 was 1.097 accidents/year) its average in subsequent years was less than K but larger than that mean. Conversely, if K was less than the mean for the population, in subsequent years it increased to a value larger than K but less than that mean. This 'shrinking' or 'regressing' to the mean is precisely what is accomplished by the EB estimator in Equation 11.1. To see that, rewrite the equation as $E\{\kappa|K\}=K+\alpha(E\{\kappa\}-K)$. Thus, if an entity recorded K accidents we estimate its κ to be 'K'+ 'correction'. The 'correction' is positive when $K<E\{\kappa\}$ and negative when $K>E\{\kappa\}$.

It remains to obtain the expression for $VAR\{\kappa|K\}$. I show in the derivations that when α is given by equation 11.2 then,

$$VAR\{\kappa|K\} = (1-\alpha)E\{\kappa|K\} \qquad \ldots 11.3$$

Had we estimated κ in the usual way, using only K for this purpose (and assuming that K is Poisson distributed) the variance of the estimate would be $E\{\kappa|K\}$. If we use both clues to safety, and estimate using Equations 11.1 and 11.2, the variance is given by Equation 11.3. Since $0 \le \alpha \le 1$, $VAR\{\kappa|K\}$ never exceeds $E\{\kappa|K\}$ and is almost always smaller. In other words, the variance of the estimate when both clues are used never exceeds, and is almost always smaller, than the variance of the estimate based on the count of accidents alone.

Numerical Example 11.3. Precision of $\hat{\kappa}$ for grade crossing in Numerical Example 11.2.

Consider the highway-rail grade crossing from Numerical Example 11.2. By Equation 11.3, $V\hat{A}R\{\hat{\kappa}|2\}=0.19\times0.48=0.09$ [accidents]^2. Thus, $\hat{\kappa}=0.48$ accidents in five years, as obtained earlier, and the standard deviation of this estimate $\hat{\sigma}\{\hat{\kappa}\}=\sqrt{0.09}=0.30$ accidents in five years.

The benefit of using both clues to safety is now obvious. Had only the accident count been used, the estimate of κ would have been two accidents in five years and the standard deviation of that estimate would be estimated at $\sqrt{2}=1.4$ accidents in five years. Using the EB method and the added information about the reference population the estimate of κ is 0.48 accidents in five years and the standard deviation of this estimate is approximately 0.30 accidents in five years. This accomplishes the remaining part of what we set out to do. Information about $E\{\kappa\}$, $VAR\{\kappa\}$ and K, has been used to produce the knowledge of $VAR\{\kappa|K\}$.

So far I have shown how to estimate the $E\{\kappa|K\}$ and the $VAR\{\kappa|K\}$ using both clues to safety. It is sometimes useful to have a model of the entire probability distribution function of $\kappa|K$.To set out in this direction one has to **assume** that the probability distribution of κ in the reference population can be adequately described. Here I will assume that the distribution of κ's in the entities of the reference population can be described by a Gamma probability density function[1]

---

[1] A Gamma probability density function can often adequately describe random variables which take on only non-negative values and whose density function is characterized by a single peak. One can not know whether this is a good assumption for a specific population of entities. However, it can be shown that **if** the accident count for an entity is Poisson distributed, **and if** in a population of entities the κ's are Gamma distributed, **then** the distribution of accident counts in the population of entities obeys the negative binomial probability distribution. Whether in a specific case the accidents counts fit the negative binomial distribution can be checked against data. I have done so for many different data sets with good results. While this is not a proof that the κ's are indeed

to be denoted by g(κ). The function g(κ) has two parameters, 'a' and 'b', defined in terms of its mean and variance by:

$$a = \frac{E\{\kappa\}}{VAR\{\kappa\}}, \quad b = \frac{(E\{\kappa\})^2}{VAR\{\kappa\}} \qquad \text{...11.4}$$

Thus, if estimates of E{κ} and VAR{κ} are available, estimates of the parameters 'a' and 'b' can be computed. Using Bayes theorem I show in the derivations that the probability density function of κ|K is also Gamma:

$$g(\kappa|K) = \frac{(1+a)^{K+b}\ \kappa^{K+b-1}\ e^{-\kappa(1+a)}}{\Gamma(K+b)} \qquad \text{. . . 11.5}$$

and that its mean and variance are:

$$E\{\kappa|K\} = \frac{K+b}{1+a}, \quad VAR\{\kappa|K\} = \frac{K+b}{(1+a)^2} \qquad \text{...11.6}$$

Substituting for 'a' and 'b' in the expressions from Equation 11.4 into Equation 11.6, we obtain, once again, Equations 11.1 and 11.3. In this manner these important results have been obtained in two different ways. Once, as a maximum precision estimator when two independent measurements of the same quantity but of different precision are combined; and once using Bayes theorem and assuming that in the reference population the κ's have a known Gamma distribution. (This is a reflection of a more general property of Bayes estimators.)

In this section I have shown how to estimate the κ of an entity using both clues. The estimator is a linear combination of the accident count and the mean of the reference population (Equation 11.1). The proportions in which the two are combined are given by Equation 11.2. The variance of this estimate is given by Equation 11.3. These results do not rest on any assumption about the distribution of κ in the reference population. If in the reference population the κ's are Gamma distributed, then in the subpopulation of entities which all recorded K accidents the κ's are also Gamma distributed as shown by Equation 11.5. To make use of these results one has to have estimates of E{κ} and VAR{κ}. How these can be obtained is discussed in the next section.

Gamma distributed, for the many data sets examined the empirical evidence was found to be consistent with the Gamma assumption.

Derivations.

Preliminaries: How to join two measurements of differing precision.

To prepare the ground, consider the following general problem. Let X and Y be two independent random variables with variances VAR{X}, VAR{Y} and α a constant. Define a new random variable Z by $Z=\alpha X+(1-\alpha)Y$. With this definition,

$$VAR\{Z\}=\alpha^2 VAR\{X\}+(1-\alpha)^2 VAR\{Y\} \qquad \ldots 11.7$$

That α which will make VAR{Z} smallest will satisfy

$$\frac{dVAR\{Z\}}{d\alpha}=2\alpha VAR\{X\}-2(1-\alpha)VAR\{Y\}=0 \qquad \ldots 11.8$$

From here it follows that VAR{Z} will be smallest when

$$\alpha=\frac{VAR\{Y\}}{VAR\{X\}+VAR\{Y\}} \qquad \ldots 11.9$$

Dividing the numerator and denominator by VAR{X}VAR{Y}, this can be rewritten as

$$\alpha=\frac{\frac{1}{VAR\{X\}}}{\frac{1}{VAR\{X\}}+\frac{1}{VAR\{Y\}}} \qquad \ldots 11.10$$

and

$$1-\alpha=\frac{\frac{1}{VAR\{Y\}}}{\frac{1}{VAR\{X\}}+\frac{1}{VAR\{Y\}}} \qquad \ldots 11.11$$

Thus, the 'weights' α and $1-\alpha$ are inversely proportional to the variances of the corresponding random variables. When the α which makes VAR{Z} smallest is used in Equation 11.7,

$$VAR\{Z\}=\frac{VAR\{X\}\,VAR\{Y\}}{VAR\{X\}+VAR\{Y\}}=\frac{1}{\frac{1}{VAR\{X\}}+\frac{1}{VAR\{Y\}}} \qquad \ldots 11.12$$

$$=\alpha VAR\{X\}=(1-\alpha)VAR\{Y\}$$

Derivation of Equations 11.2 and 11.3.

The above results can now be applied to the problem at hand. We have two clues about the κ of a certain entity - 'our κ'. The first is derived from the reference population that matches the known traits of this entity. In the reference population the κ's are not all the same, they have a mean $E\{\kappa\}$ and variance $VAR\{\kappa\}$. Using only the information about the reference population, the best estimate[1] of 'our κ' is $E\{\kappa\}$. If so, the variance of the estimate of 'our κ' is $VAR\{\kappa\}$. Thus, were we to use only the first clue, we would estimate the κ of the entity to be $E\{\kappa\}$. The second clue about κ is that our entity has recorded K accidents in that period. Were we to use only the second clue, we would estimate 'our κ' to be K. If K is Poisson distributed with a mean κ, the variance of K is also κ. Since we do not know at this point of the analysis what this κ is, it seems sensible to use $E\{\kappa\}$ for the variance of K. Thus, we have two independent estimators of the same quantity. The correspondence with the preliminary analysis is as follows:

X corresponds to $E\{\kappa\}$ and $VAR\{X\}$ to $VAR\{\kappa\}$;
Y corresponds to K and $VAR\{Y\}$ to $E\{\kappa\}$.

Since we wish to use both clues, we join the two estimators in a linear combination with 'weights' α and $(1-\alpha)$ where $0 \leq \alpha \leq 1$.

$$\hat{\kappa}=\alpha E\{\kappa\}+(1-\alpha)K \qquad \ldots 11.13$$

As is shown above, when two estimates of unequal precision are joined, the weights α and $1-\alpha$ that minimize the expected squared error of estimation are inversely proportional to the variance of the estimates. In our case, α is proportional to $1/VAR\{\kappa\}$ and $1-\alpha$ is proportional to $1/E\{\kappa\}$. Substituting now into Equation 11.10,

$$\alpha = \frac{\frac{1}{VAR\{\kappa\}}}{\frac{1}{VAR\{\kappa\}}+\frac{1}{E\{\kappa\}}} = \frac{1}{1+\frac{VAR\{\kappa\}}{E\{\kappa\}}} \qquad \ldots 11.14$$

This is Equation 11.2a.

To obtain Equation 11.3, use is made of Equation 11.12. As before replace $VAR\{X\}$ by $VAR\{\kappa\}$. At this point in the analysis we already have an estimate of κ, the mean of K. It has denoted by $E\{\kappa|K\}$ and can be used to represent the $VAR\{Y\}$ in Equation 11.12.

---

[1] The expected value is the best estimate of any random variable in the sense that it makes the mean squared error of estimation smallest.

The inconsistency in what is used for the variance of K (first E{κ} and later E{κ|K}) is a blemish of this derivation and a challenge for future inquiry. Would one estimate better is E{κ|K} was used in both cases? The merit of this derivation is in that it does not require any assumptions about the distribution of the κ's in the reference population and agrees with the results of the derivation in which one assumes that the κ's are Gamma distributed.

Derivation of Equations 11.5 and 11.6.

Analysis here rests on two assumptions. One assumption is that the distribution of κ's in the reference population can be described by a Gamma probability density function to be denoted by g(κ). That is, for $\kappa \geq 0$,

$$g(\kappa) = \frac{a^{b} \kappa^{b-1} e^{-a\kappa}}{\Gamma(b)} \qquad \ldots 11.15$$

The parameters 'a' and 'b' are related to E{κ} and VAR{κ} as follows:

$$E\{\kappa\} = \frac{b}{a}\ , \quad VAR\{\kappa\} = \frac{b}{a^2} \qquad \ldots 11.16$$

$$a = \frac{E\{\kappa\}}{VAR\{\kappa\}}\ , \quad b = \frac{(E\{\kappa\})^2}{VAR\{\kappa\}} \qquad \ldots 11.17$$

The notation Γ(b) stands for the value of the gamma function[1] with the argument 'b'. Thus, if estimates of E{κ} and VAR{κ} are available, estimates of the parameters 'a' and 'b' can be computed. The other assumption is that the count of accidents K is Poisson distributed. Thus,

$$P(K|\kappa) = \frac{\kappa^{K} e^{-\kappa}}{K!} \qquad \ldots 11.18$$

According to Bayes theorem for probability distributions,

$$g(\kappa|K) = (constant)\, P(K|\kappa)\, g(\kappa) \qquad \ldots 11.19$$

Using the expressions for P(K|κ) (Equation 11.18) and g(κ) (Equation 11.15) in Equation 11.19, Equation 11.20 obtains. This too is a Gamma probability density function similar to Equation 11.15 except that 'b' has been replaced by 'K+b' and 'a' by '1+a'.

---

[1] For positive b, Γ(b) is the value of integral of $e^{-t}\, t^{b-1}$ dt over all non-negative values of t, which for integer values of b is (b-1)!.

$$g(\kappa|K) = (constant_1)\ \frac{\kappa^K e^{-\kappa}}{K!}\ \frac{a^b \kappa^{b-1} e^{-a\kappa}}{\Gamma(b)}$$
$$= (constant_2)\ \kappa^{K+b-1} e^{-\kappa(1+a)} \qquad \ldots 11.20$$

After integration over all $\kappa \geq 0$, the normalizing $constant_2$ turns out to be $(1+a)^{K+b}/\Gamma(K+b)$. Thus, we obtain Equation 11.5

$$g(\kappa|K) = \frac{(1+a)^{K+b}\ \kappa^{K+b-1} e^{-\kappa(1+a)}}{\Gamma(K+b)} \qquad \ldots 11.5$$

In analogy with Equation 11.16, the mean and the variance are now:

$$E\{\kappa|K\} = \frac{K+b}{1+a}, \quad VAR\{\kappa|K\} = \frac{K+b}{(1+a)^2} \qquad \ldots 11.6$$

11.5 How to estimate E{κ} and VAR{κ}

The pivotal concept of the EB method is that of a reference population. Each entity of the reference population was said to have its own expected accident count κ. The EB method uses information about the mean and the variance of the κ's in the reference population (see Equations 11.1, 11.2 and 11.4). Therefore, methods for estimating $E\{\kappa\}$ and $VAR\{\kappa\}$ have to be provided. Two such methods are discussed in this section. The first is simple, but can be used only rarely. The second is more widely applicable, but involves new concepts and therefore requires more detailed explanation. Common to both methods are the two equations below. Their derivation is given at the end of the section.

$$E\{K\}=E\{\kappa\} \qquad \ldots 11.21$$
$$VAR\{K\}=E\{\kappa\}+VAR\{\kappa\} \qquad \ldots 11.22$$

In words, the expected value of the accident counts in the reference population is the same as the expected value of the κ's in the reference population. The variance of the accident counts in the reference population is the sum of the expected value of the κ's in the reference population and their variance. Equations 11.21 and 11.22 suggest a simple method for the estimation of $E\{\kappa\}$ and $VAR\{\kappa\}$.

a. The Method of Sample Moments

Consider a reference population of n entities of which n(K) entities have recorded K=0, 1, 2, . . . accidents during a specified period. With this notation, the sample mean and the sample variance are

$$\bar{K}=\sum Kn(K)/n \qquad \ldots 11.23$$

$$s^2=\sum (K-\bar{K})^2 n(K)/n \qquad \ldots 11.24$$

The summation is over all values of K. As n increases, $\bar{K}$ approaches E{K} and s^2 approaches VAR{K}. Thus, replacing E{K} by $\bar{K}$, and VAR{K} by s^2 in Equations 11.21 and 11.22 we obtain,

$$\hat{E}\{\kappa\} = \bar{K} \qquad \ldots 11.25$$

$$\widehat{VAR}\{\kappa\} = s^2 - \bar{K} \qquad \ldots 11.26$$

The larger the reference population, the more accurate are these estimates. Because the estimators of E{κ} and VAR{κ} are based on $\bar{K}$ and s^2, this approach to the estimation will be called the **Method of Sample Moments**.

To illustrate its use I will introduce two numerical examples. For the first example I return to the data of Table 11.2 (in Section 11.2). This data has been used earlier to show that intersections that in 1974 recorded K accidents did not record on the average K accidents in 1975. Here I will use the same data to show how E{κ} and VAR{κ} are estimated by the Method of Sample Moments. I will also show how to use this information to estimate the κ of an intersection that recorded K accidents.

In the second numerical example I use data about highway-rail grade crossings. I will use this example later to show what problems beset the Method of Sample Moments. The same example will then be continued when the alternative method of estimation will be discussed.

Numerical Example 11.4. Method of Sample Moments applied to stop controlled intersections in San Francisco.

As in Table 11.2, column 1 of Table 11.4 gives the number of intersections (n(K)) for which the count of accidents in 1974 was K=0,1,2,. . . . On this basis, the entries of Columns 3 and 4 are computed. Their sums are needed to compute the sample mean and sample variance in Equations 11.23 and 11.24.

Table 11.4. Accident counts at 1142 intersections 1974/1975

1	2	3	4
n(K)	K	K×n(K)	$(K-\bar{K})^2\times n(K)$
553	0	0	665.73
296	1	296	2.80
144	2	288	117.37
65	3	195	235.34
31	4	124	261.21
21	5	105	319.87
9	6	54	216.34
13	7	91	452.96
5	8	40	238.24
2	9	18	124.91
2	13	26	283.35
1	16	16	222.09
1142		1253	3140.21

Using the results from Table 11.4, $\bar{K}$=1253/1142=1.097 accidents/year and s^2=3140.21/1142 =2.750. From here, $\hat{VAR}\{\kappa\}=s^2-\bar{K}$=2.750-1.097=1.653. These results can now be used to estimate $\hat{\alpha}$ by Equation 11.2a ($\hat{\alpha}$=1/(1+1.653/1.097)=0.399). Equations 11.1 and 11.3 are used to estimate E{κ|K} and VAR{κ|K}. Thus, e.g., for an intersection which in 1974 recorded 2 accidents (K=2), $\hat{E}\{\kappa|K\}$=0.399×1.097+(1-0.399)×2=1.64 accidents/year and $[\hat{VAR}\{\kappa|K\}]^{1/2}=[(1-0.399)\times 1.64]^{1/2}$=0.99 accidents/year. The actual average number of accidents in 1975 at those 144 intersections that in 1974 recorded 2 accidents was 1.53. The standard deviation of this average is estimated to be 0.99/√144=0.08. Thus, the correspondence between the EB estimate and the observed average is, in this case, satisfactory. A more detailed examination of the correspondence between the estimate produced by the EB method and the actual average accident counts will be given in the next section.

Numerical Example 11.5. Method of Sample Moments applied to highway-rail grade crossings.

Consider a reference population of rail-highway grade crossings the defining traits of which are: {urban area, 1 track, 1 to 2 trains per day, 0 to 1000 vehicles per day, signed by crossbucks}. In 1980 there were 9939 such crossings in the U.S.A. of which 9770 had 0 accidents, 160 had 1 accident, 8 had 2 accidents and 1 had 3 accidents. From the computations in the tableau we find that $\bar{K}$=179/9939=0.0180 [accidents/year], s^2=197.78/9939=0.0199 [accidents/year]^2 and therefore the estimates are $\hat{E}\{\kappa\}$=0.0180 [accidents/year] and $V\hat{A}R\{\kappa\}$=0.0199-0.0180=0.0019 [accidents/year]^2. From here, $\hat{\alpha}$= 1/(1+0.0019/0.0180) =0.905. Using this, estimates of $E\{\kappa|K\}$ were calculated in the rightmost column of the tableau. Thus, crossings that had no accidents in that year are estimated to have κ=0.016 accidents/year while the crossings that has recorded 3 accidents in that year are estimated to have κ=0.301 accidents/year.

n(K)	K	K×n(K)	$(K-\bar{K})^2\times n(K)$	$\hat{E}\{\kappa\|K\}$
9770	0	0	3.17	0.016
160	1	160	154.29	0.111
8	2	16	31.43	0.206
1	3	3	8.89	0.301
9939		179	197.78	

The attraction of the Method of Sample Moments is that its validity rests on a single assumption namely, that if the κ of an entity did not change with time, the occurrence of accidents on it would be well described by the Poisson probability law. However, grave practical difficulties arise. To have a reference population, its entities have to match the traits of the entity for which κ is estimated. Suppose then that we are interested in the safety of a four legged intersection in San Francisco with STOP signs on the minor approaches, and with 540 and 2300 vehicles per day on the minor and major approaches. The problem is that, the estimates in Numerical Example 11.4 are not sufficiently specific. They are for a reference population with similar traits (layout, traffic control) but the 'traffic flow' traits are not represented. Were we to attempt estimation of $\bar{K}$ and s^2 using only those intersections that approximately match the traffic flows of the entity of interest, there would be so few matching intersections in San Francisco, that the estimates would be so imprecise as to be useless. Only rarely can one find a sufficiently large data set to allow a reasonable matching of traits, while still providing adequately accurate estimation of $\bar{K}$ and s^2 (and therefore of E{κ} and VAR{κ}). Even with very large data sets, one cannot find adequate reference populations when entities are described by several traits, or when some traits that are continuous in nature (such as traffic flow). Thus, in Numerical Example 11.5, with more than 200,000 grade crossings in the USA, use had to be made of the broad categories '1 to 2 trains per day' and '0 to 1000 vehicles per day', even though the corresponding traits of the entity in Numerical Examples 11.2 and 11.3 are '2 trains per day' and '550 vehicles per day'. It is for this reason that the estimates of E{κ} and VAR{κ} of Numerical Example 11.5 were not used in the previous numerical examples. The Multivariate Regression Method discussed below is an attempt to obviate these difficulties.

b. The Multivariate Regression Method

It is often necessary to estimate the safety of entities for which a sizeable reference population does not or cannot exist. To illustrate, if it is known that Mr. Smith (the 22-year-old Ontario driver who has been introduced to the reader in Section 11.3) has driven in 1984 about 13500 kilometers and has been convicted three times for speeding and once for impaired driving, conceptually there still is a reference population. Its defining traits are:

> {Ontario driver in 1984; male; 22 years old; 13500 km/year; three speeding convictions; one impaired driving conviction}.

However, although there are more than six million drivers in Ontario, the number of drivers who match all these traits is too small to allow reliable estimation of E{κ} and VAR{κ} by the Method of Sample Moments. In the limit, when traits are many or are continuous in nature (as are age, or average traffic flow), a real reference population does not exist. In this case we can only ask what **would be** the mean and the variance of the κ **if** a large reference population did exist. The concept of a reference population is still clearly defined but it is now an **imagined** one. Since no data about it can exist, the Method of Sample Moments cannot possibly be used. It appears, however, that even here it is still feasible to estimate E{κ} and VAR{κ}. Not only is estimation feasible, the resulting estimates can now be tailored precisely to the traits of the entity the κ of which is to be estimated. The idea for doing so is explained below. It is the same idea as that on which much of multivariate statistical regression analysis rests.

It is common to estimate the safety of entities with specified traits by multivariate statistical models. Thus, e.g., there are 'multivariate models' to estimate the safety of a grade crossing as a function of the train and car traffic, the location of the crossing, its geometric characteristics, and the type of warning device used. Similarly, there exist multivariate models that estimate the safety of a driver as a function of age, gender and number of convictions for offenses associated with the use of the road. Multivariate models exist also for estimating the safety of intersections and roads. The purpose of this section is to elucidate the relationship between such multivariate models, imagined reference populations, and the two moments, $E\{\kappa\}$ and $VAR\{\kappa\}$.

Imagine a population of entities with identical measured traits. In such an imagined reference population, as in any real one, κ would still vary from entity to entity. To see why, consider a group of Ontario drivers who, just like Mr. Smith, are male, 22 years old, have driven 13500 km and got three speeding and one impaired driving conviction in 1984. One should still expect them to be different from each other and from Mr. Smith in important but unmeasured respects - education, personality, where they drive, what cars they use, and so on. Therefore drivers in this imagined reference population should also be expected to have κ's that are not all the same. For this reason, even for such a homogeneous population, one can still make meaningful statements about $E\{\kappa\}$ and $VAR\{\kappa\}$.

When a multivariate model is being fitted to accident data, it is to estimate the $E\{\kappa\}$ as a function of variables called 'covariates'. This activity is based on the belief that $E\{\kappa\}$ depends on the covariates in some systematic way which can be captured by an equation or a set of equations - a model. These **covariates** are precisely what has so far been called **traits**. Thus, each set of covariate values defines an imagined reference population.

When a multivariate model is used to estimate the κ of some specific entity, the reasoning mimics that used earlier in Section 11.3. To illustrate, assume that the aim is to estimate the κ of Mr. Smith. The reasoning goes as follows:

> We have an estimate of $E\{\kappa\}$ obtained from the multivariate model for that imagined reference population of Ontario drivers that matches exactly Smith's traits. Because Smith is indistinguishable from others in this imagined reference population, the best estimate for Smith's κ is what $E\{\kappa\}$ is estimated to be by the multivariate model for this imagined reference population.

From here on we will see the multivariate model as providing estimates of $E\{\kappa\}$ for a continuum of imagined reference populations. This interpretation of what a multivariate models represents logically entails also a specific view of the 'residuals' as shown in Figure 11.3.

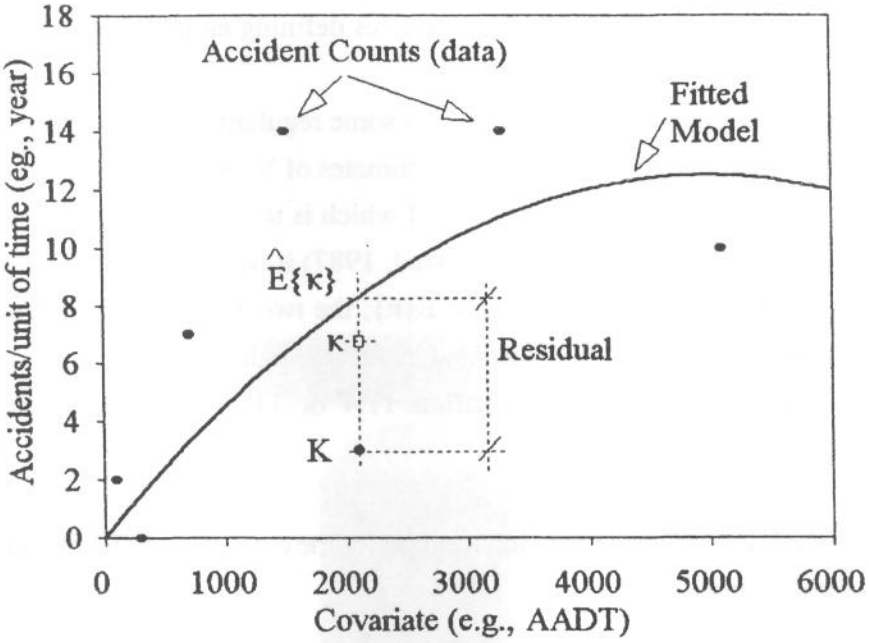


Figure 11.3. Fitted values and residuals.

As always, a **residual** is the difference between an accident count for some specific entity which served as 'datum' for model fitting and the estimate of E{κ} calculated from the fitted model equation. This residual can be thought of as consisting of two parts. One part is the difference between the (unknown) κ of the specific entity which happened to provide the observed datum and the count of accidents on it. This part is usually taken to be the difference between a Poisson-distributed count (K) and its mean (κ). The other part is the difference between the κ of the specific entity which served as a datum and between the estimate of E{κ} for the imagined reference population to which the entity belongs. In other words, K is a realization of a Poisson random variable the mean of which is κ, while κ is a realization from an imagined reference population in which κ's have a distribution with E{κ} and VAR{κ} as central moments.[1]

Consider now one of the entities that were used to estimate the parameters of the multivariate model. Using its traits and the multivariate model equation, we can now estimate the E{κ} of the reference population to which the entity belongs. The squared difference between the accident count of this entity and the estimate of E{κ} estimates the VAR{K} in the imagined reference population. Therefore, based on the relationship in Equation 11.22, the difference [squared residual - estimate of E{κ}] can be used to estimate the VAR{κ} for the imagined reference population to which this

[1] In effect, the residuals are viewed as coming from a family of compound Poisson distributions. For compound Poisson distributions VAR{K}=VAR{κ}+E{κ} as given earlier in Equation 11.22.

entity belongs. Thus, for each entity used in the multivariate modeling, there is an estimate of VAR{κ} which is associated with a set of covariates defining an imagined reference population.

Just as E{κ} is related to the covariates by some regularities that can be captured by a model equation, so might VAR{κ} be. Therefore, the estimates of VAR{κ} for every datum point can also be subjected to a multivariate analysis the aim of which is to express VAR{κ} as a function of the covariates. In previous work (Hauer and Persaud, 1987) it has been found that when estimates of VAR{κ} are plotted against the estimates of E{κ}, the two are systematically associated. Their relationship can often be adequately represented by the quadratic function[1] $VAR\{\kappa\}=[E\{\kappa\}]^2/b$. The parameter 'b' appeared earlier (see Equation 11.4 or 11.17) where the underlying Gamma distribution has been specified.

In sum, the Multivariate Regression Method for the estimation of E{κ} and VAR{κ} consists of the following activities:

1. Accident counts and measured traits (covariate values) for 'n' entities serve as data. Fit a multivariate model to the data. The model yields estimates $\hat{E}\{\kappa\}$ of E{κ} as a function of the specific values that the covariates assume.

2. For each entity calculate the residual - the difference between the accident count and the fitted value $\hat{E}\{\kappa\}$. Using this, compute for each entity the difference [squared residual - $\hat{E}\{\kappa\}$]. This difference is the $V\hat{A}R\{\kappa\}$ for this entity. Thus, at the end of this step we have a value $V\hat{A}R\{\kappa\}$ for each of the n entities that have been used in model fitting.

3. Fit a multivariate model linking the n values of $V\hat{A}R\{\kappa\}$ and the covariate values of the corresponding entities. At times it is possible to abridge the process by fitting a model that makes the $V\hat{A}R\{\kappa\}$ a simple function of the $\hat{E}\{\kappa\}$ which already has been expressed as a function of the covariates of the n entities.

How the results of the Multivariate Regression Method for the estimation of E{κ} and VAR{κ} might be used, is illustrated Numerical Example 11.6.

---

[1] There is no reason to believe that a relationship of this kind holds in every data set. When, in view of the data at hand, a quadratic relationship (or any other relationship making VAR{κ} a function of E{κ}) is justified, this is a particularly simple way to link VAR{κ} with the covariates which are used to estimate E{κ}. Only one added parameter, the coefficient of proportionality needs to be estimated. Theoretical considerations and the empirical support for the choice of this simple quadratic relationship are given in Hauer (1989b).

Numerical Example 11.6. The Multivariate Regression Method applied to grade crossings.

Consider again the highway-rail grade crossing of the preceding numerical examples. Using a multivariate model developed earlier (Hauer and Persaud, 1987), $\hat{E}\{\kappa\}$=constant× $(\text{cars/day})^{\beta}$ ×$(\text{trains/day})^{\gamma}$×$(\text{trains/day})^{\delta\times\ln(\text{trains/day})}$. For the crossing in Numerical Example 11.2 (urban, single track, two trains per day, 550 cars per day, crossbucks), $\hat{E}\{\kappa\}$=0.0239 accidents/year. In this, constant= 0.000954, β=0.405, γ=1.039 and δ=-0.115, which are parameter estimates specific to the traits: {urban, single track, crossbucks}. In the same multivariate model and set of traits we found that $V\hat{A}R\{\kappa\}=\hat{E}^2\{\kappa\}/0.52= 0.0239^2/0.52=0.0011$ [accidents/year]^2. These are the estimates of $E\{\kappa\}$ and $VAR\{\kappa\}$ used earlier in Numerical Example 11.2.

The advantages of the Multivariate Regression Method are two and they are linked. First, it provides estimates of $E\{\kappa\}$ and $VAR\{\kappa\}$ for a reference population that matches the traits of the entity of interest **exactly**. Traits that are continuous in nature cease to be a problem. Of course, only those traits can be matched that are included in the multivariate model. Second, one does not need to have a large reference population for any particular combination of traits. Thus, there are practical ways to implement the Empirical Bayes method of safety estimation the elements of which were described in this and the preceding section. It is time to put it to the test.

Derivation.

Equations 11.21 and 11.22.

In Section 10.1 I have discussed (in a general notation) the following problem. A random variable K can take on values 0, 1, 2, ... There exist 'n' distributions of K with means and variances (κ_1,v_1), (κ_2,v_2), . . . , (κ_j,v_j), . . . , (κ_n,v_n). A 'trial' is said to consist of taking one observation of K from each of the n distributions. We were interested in the mean and variance of the K's in a sequence of such trials and in their relationship with the (κ_1,v_1), (κ_2,v_2), . . . , (κ_j,v_j), . . . , (κ_n,v_n). I have shown there that

$$E\{K\} = \frac{\sum_{j=1}^{j=n}\kappa_j}{n} \doteq \bar{\kappa} \qquad \text{and} \qquad VAR\{K\} = \frac{\sum_{j=1}^{j=n} v_j}{n} + \frac{(\sum_{j=1}^{j=n}\kappa_j^2) - n\bar{\kappa}^2}{n} \doteq \bar{v} + s_\kappa^2$$

Consider now a reference population where K denotes accident counts that are Poisson distributed. In the reference population entities have the expected accident counts κ_1, κ_2, . . . , κ_i,

$\ldots,\kappa_n$ and, for all i, $v_i=\kappa_i$. Thus $\bar{\kappa}=\bar{v}$. For large 'n', $\bar{\kappa}\rightarrow E\{\kappa\}$ and $s^2_\kappa \rightarrow VAR\{\kappa\}$. In this manner, when n is large, in the equations above $E\{K\}\rightarrow E\{\kappa\}$ and $VAR\{K\}\rightarrow E\{\kappa\}+VAR\{\kappa\}$.

A simple way to obtain the same result is to use two propositions about conditional moments from probability theory[1]. For jointly distributed random variables X and Y,

$$E\{X\}=E\{E\{X|Y\}\} \quad \ldots a$$

and

$$VAR\{X\}=E\{VAR\{X|Y\}\}+VAR\{E\{X|Y\}\} \quad \ldots b$$

In our case K (the count of accidents on an entity of the reference population) stands for X, and κ (the expected value of K) stands for Y. By definition, $E\{K|\kappa\}=\kappa$. Using this in equation 'a', we get $E\{K\}=E\{\kappa\}$. Since K is Poisson distributed, $VAR\{K|\kappa\}=\kappa$. Using this in equation 'b' we find $VAR\{K\}=E\{\kappa\}+VAR\{\kappa\}$.

11.6 THE PROOF OF THE PUDDING

In observational studies one may not assume that the count of accidents in the period before treatment had nothing to do with the reasons why a treatment has been applied (or withheld). Therefore, in observational studies, the count of accidents (K) in some 'before' period is a biased estimate of the expected accident count in that period (κ). The bias has been called the 'selection bias' or the 'regression-to-mean' (RTM) bias. In Section 11.2 I have discussed the logical reasons for the RTM bias and have shown its existence using data. To obviate the RTM problem I have suggested that safety estimation be based on mixing two clues to safety; those that derive from the knowledge of the traits of the entity and the accident counts of its reference population, and those that come from the accident counts of the entity the safety of which is being estimated. This approach to estimation as described in Sections 11.3 to 11.5 is known as the Empirical Bayes (EB) approach for safety estimation. The claim is that the EB method will cleanse safety estimation of the RTM bias. What grounds do I have for this claim?

First, it can be said that the conceptual frame of the EB method fits the reality of observational studies. The essence of the EB idea is that if an entity recorded K accidents of some kind, we estimate its κ by estimating $E\{\kappa|K\}$, the mean of the κ's in the subpopulation of the reference population. We think of a two-stage selection process. In the first stage we identify from

---

[1] See, e.g., S. Ross, A First Course in Probability, Third Edition, MacMillan Publishing Co., 1988, Propositions 6.1 and 6.2, pp. 285-292.

the set of all possible entities those that have specified traits - the imagined reference population. Next from the entities of the imagined reference population we select yet another subset, those entities that recorded K accidents. Thus, the act of selection which gives rise to the RTM bias, is an integral part of the logical framework of the EB method. This approach to estimation explicitly recognizes the fact that entities that recorded K accidents have a mean that is different from the mean characterizing entities that recorded L accidents ($K \neq L$).

Second, it is clear from Equation 11.1 that whether K is larger than $E\{\kappa\}$ or smaller, the estimator for $E\{\kappa|K\}$ is always between K and $E\{\kappa\}$. Thus, at least qualitatively, $E\{\kappa|K\}$ does what the logic of RTM predicts and what data show (see, e.g., Figure 11.1). Namely, it shifts the estimate for K in the direction of the population mean. It remains to show that what is logically sound and leads to qualitatively correct conclusions, is also born out by data.

The existence and extent of the RTM bias have been illustrated earlier using data about 1142 intersections in San Francisco (see Tables 11.2, 11.3 and Figure 11.1). In Figure 11.1, e.g., I show how the 1974 accident count differs from the average in the subsequent years. We can now juxtapose the EB estimate with the average in the subsequent year and see how the two compare. From Numerical Example 11.4, $\bar{K}$=1.097 and $\hat{\alpha}$=0.399. With this, using Equations 11.1 and 11.2, the entries of columns 3 of Table 11.5 have been calculated. Column 4 is the average accident count recorded in 1975 as shown in Table 11.2.

Table 11.5. Juxtaposition of EB estimates for 1974 and observed averages in 1975.

1	2	3	4
n(K)	K	$\hat{\kappa}=\hat{E}\{\kappa\|K\}$	avg(K)
553	0	0.44	0.54
296	1	1.04	0.97
144	2	1.64	1.53
65	3	2.24	1.97
31	4	2.84	2.10
21	5	3.44	3.24
9	6	4.04	5.67
13	7	4.64	4.69
5	8	5.25	3.80
2	9	5.85	6.50

One should be comparing the estimate in column 3 with the average in column 4. These results are shown in Figure 11.4. The ordinate of each full square is the EB estimate of $\hat{\kappa}$

(=$\hat{E}\{\kappa|K\}$) for 1974. It is plotted against avg(K) for 1975 on the abscissa. Were the two equal, the squares would align themselves on the straight line shown. The first six squares do so quite well. However, the squares in the right half of the figure fluctuate around the straight line. The reason is that in this case avg(K) has a large standard deviation, since it is based on a small n(K), and so does $\hat{E}\{\kappa|K\}$.

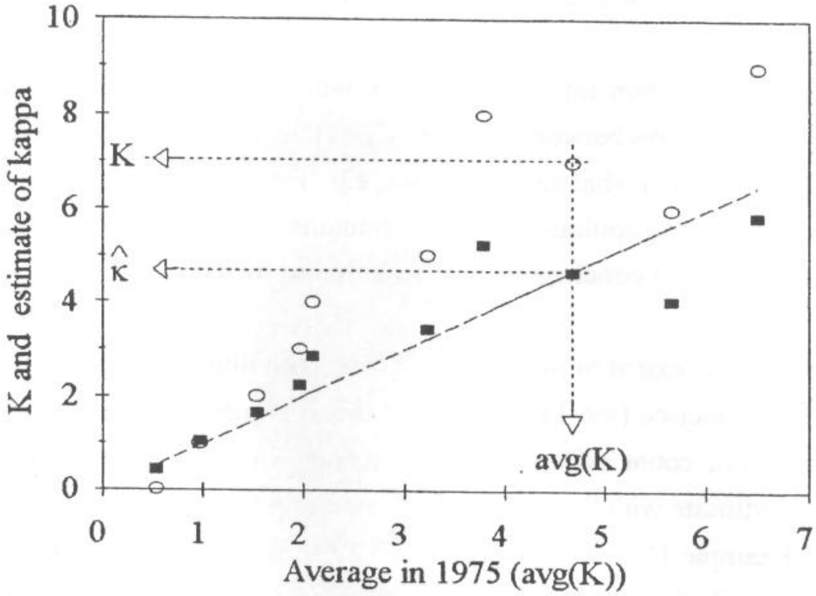


Figure 11.4. Comparing the EB estimate (for 1974) and K (for 1974) avg(K) (for 1975).

The ordinate of the empty circles is K in 1974. As noted earlier, the difference between accident count in 1974 and the average in 1975 (avg(K)) is systematic. When K is less than the mean for the population (which is about 1.1 accidents/year), the circle is below the straight line and vice versa. One may conclude that for this data the EB estimates correspond better to the average in 1975 than do the 1974 accident counts.

There are several factors which account for the difference between $\hat{\kappa}$ and avg(K). First, I already noted the statistical uncertainty in avg(K). Second, the $\hat{\kappa}$ measured on the vertical axis also has a similar variance. Third, we are comparing estimates for 1974 with averages in 1975. Fourth, the method rests on the assumption that K is Poisson distributed which is merely some useful approximation to reality. Considering all these reasons for the difference between $\hat{\kappa}$ and avg(K), one may deem the performance of the estimator in this data set to be quite satisfactory.

We have examined in a similar manner a large number of diverse data sets (Hauer and Persaud, 1983 and Persaud and Hauer, 1984). In each case the EB method produced estimates that are free of the RTM bias. Therefore, I conclude, that the EB method is a satisfactory tool for safety estimation that is free of the RTM bias.

11.7 TWO CASE STUDIES

In the early 1980s it seemed that ridding results of Before-After studies of the regression-to-mean bias was the main task at hand. It now may seem peculiar that what could be considered 'well-known' in other disciplines did not percolate into the traffic safety literature till then. Thus, it was only then that I came to the (belated) realization, that when entities are treated because they have an unusually bad accident record, a Before-After study will generally exaggerate the effectiveness of treatments (see, e.g., Hauer 1980a). Since the RTM bias was present in the accounts of many Before-After studies in the professional literature, the aggregate result was (and perhaps continues to be) an overly optimistic view of how safety can be improved by various treatments. Against this background, it seemed natural to try to redeem results of Before-After studies by attempting to cleanse them of the RTM bias (see, e.g., Hauer, 1980b, Abbess et al., 1981, Hauer and Persaud, 1983). The two cases reviewed below serve as an illustration of early applications of the EB method.

a. Conversion from 'two-way' to 'four-way' stop control in San Francisco

This case study describes an early application of the Method of Sample Moments (Hauer et al, 1984). In 1973 the California Office of Traffic Safety provided a grant to the San Francisco Department of Public Works to conduct a study of that city's high accident intersections. The result was a report published in 1974, documenting the change in accident experience attributed to the installation of four-way-stop signs and of some other traffic control devices. According to a one year before-and-after comparison of 49 intersections converted from two-way to four-way stop control, in the period 1969-1972, the reduction in total accidents appeared to be 71%. Right-angle and injury accidents were reported to be reduced by 88% and 81% respectively. These results later found their way into the published professional literature (Short et al., 1982). As noted, the programme in San Francisco was aimed at rectifying high-accident-locations and, since the original analysis did not account for the RTM, the reported results were likely to be exaggerated.

We tried to obtain the additional data needed to remove the bias. The additional data needed are about an appropriate reference population - here, intersections with two-way stop control. The City of San Francisco was approached in 1983. Although the one year 'before' and 'after' accident data for the 49 intersections converted from two-way to four-way-stop control were still on file, the accident data covering the study period for all other intersections in San Francisco were no longer available. The closest dates for which data were obtainable were the four years running from 1974 through 1977. This was supplied to us in hard copy along with a list of all intersections controlled by two-way stop signs. We then extracted from the hard-copy all accident data pertaining to 1142 intersections controlled by two-way stop signs. (Some of this data was used in the numerical examples and illustrations earlier in this chapter.) Next we calculated the sample mean and sample variance for each year in the 1974-1977 period. But, since the sites were converted to four-way stop control in the years 1969-1972, extrapolation from 1974-1977 back into 1969-1972 was required. For 'right-angle accidents', the extrapolated values of $\bar{K}$ and s^2, and of the $\hat{\alpha}$ based on these, are listed in Table 11.6.

Table 11.6. Sample moments and $\hat{\alpha}$ for right-angle accidents.

	1969	1970	1971	1972
$\bar{K}$	0.966	0.925	0.884	0.843
s^2	2.702	2.503	2.305	2.106
s^2-$\bar{K}$	1.737	1.578	1.421	1.263
$\hat{\alpha}$	0.357	0.370	0.383	0.400

Using $\bar{K}$ for $E\{\kappa\}$ and $\hat{\alpha}$ for α in Equation 11.1 one can compute $\hat{E}\{\kappa|K\}$ for any K. Thus, e.g., if K in 1969 was 4, then $\hat{E}\{\kappa|K\}$ for 1969 is $0.357\times0.966+(1-0.357)\times4=2.92$. That is, if an intersection recorded 4 right-angle accidents in 1969, we would estimate its κ in that year to be 2.92 right-angle accidents/year. Corresponding values for other years and K's are given in Table 11.7.

Table 11.7. Estimates of $E\{\kappa|K\}$ in the 'before' years.

K	1969	1970	1971	1972
0	0.35	0.34	0.34	0.34
1	0.99	0.97	0.96	0.94
2	1.63	1.60	1.57	1.54
3	2.27	2.23	2.19	2.14
4	2.92	2.86	2.81	2.74
5	3.56	3.49	3.42	3.34
6	4.20	4.12	4.04	3.94
7	4.84	4.75	4.65	4.54
8	5.49	5.39	5.27	5.14

Among the 49 intersections, there were 3 intersections that had 4 right-angle accidents in 1969 and got converted to four-way Stop control at the end of that year. The question is, what would their κ be in 1970, had they remained untreated, that is with two-way stop control. Since $\bar{K}$ was 0.966 in 1969 and 0.925 in 1970, our prediction was the κ in 1970 would have been $2.92\times0.925/0.966=2.79$ right-angle accidents. This is the value listed for K=4 in the 1970 column of Table 11.8. Also listed in the cell, in brackets, is the number of converted intersections. Thus three intersections had four accidents in 1970 and five intersections had no accidents in 1972. The numbers in brackets add to 49.

Table 11.8 Predictions of κ for right-angle accidents in the 'after' year.

K	1970	1971	1972	1973
0	0.33 ; (2)	0.33 ; (2)	0.32 ; (5)	
1		0.93 ; (1)	0.91; (6)	0.89 ; (1)
2	1.56 ; (2)	1.53 ; (3)	1.50 ; (5)	
3	2.18 ; (1)	2.13 ; (2)	2.09 ; (4)	2.03 ; (1)
4	2.79 ; (3)		2.67 ; (1)	
5		3.34 ; (2)	3.26 ; (2)	
6		3.94 ; (1)	3.85; (1)	3.74 ; (1)
7		4.54 ; (1)		
8	5.25 ; (2)			

Thus, had these intersections not been converted to four-way stop control, one should expect in the 49 'after' years a total of $0.33\times2+1.56\times2+2.18\times1+\ldots+2.03\times1+3.74\times1=93.05$ right-angle accidents. The number of right-angle accidents actually recorded in the 49 'after' years was 16. Thus, there seemed to be a reduction of 93-16=77 right-angle accidents/year. Note that the simple count of 'before' accidents is $0\times2+2\times2+3\times1+\ldots+3\times1+6\times1=129$. The difference between 129 and 93.05 was our estimate of the RTM bias in this case.

We have examined in a similar manner several other data sets (from Philadelphia, Toronto, Michigan) and found similarly large reductions in accidents following conversion from two-way to all-way stop control. In retrospect these studies appear simple-minded; we were focused on correcting for the RTM. However, traffic counts were used little, comparison groups were crude, and the 'before' period was fixed and short. Still, this is perhaps an instructive illustration of how the Method of Sample Moments has been applied.

b. The safety effect of warning devices at highway-rail grade crossings

We first used the Multivariate Regression Method in a project the main aim of which was to estimate the safety effect of warning devices (crossbucks signs, train activated bells & flashers, and gates) at highway-rail grade crossings (Hauer and Persaud, 1987). Because of the large data base, it was possible to apply the Method of Sample Moments as well as the Multivariate Regression Method and thus to compare results obtained by the two.

The multivariate model used was of the form: $\hat{E}\{\kappa\}$= constant×(cars/day)$^{\beta}$×(trains/day)$^{\gamma}$× (trains/day)$^{\delta \times \ln(\text{trains/day})}$ (see also Numerical Example 11.6). The model parameters (constant, β, γ and δ) were estimated using information about accident counts, AADT, train traffic, number of tracks, urban or rural setting, and warning device used for the years 1980-1984. These data were available for some 200,000 U.S. grade crossings. When the model residuals were examined, the quadratic relationship VAR$\{\kappa\}$=(E$\{\kappa\}$)^2/b was confirmed and the parameter 'b' was estimated. In addition, because the data set was very large, it was possible to calculate $\bar{K}$ and s^2 with reasonable precision for many categories of crossings. (See, e.g., Numerical Example 11.5 where the defining traits of nearly 10,000 crossings were: urban area, one track, 1-2 trains/day, 0-1000 vehicles per day, signed by crossbucks.)

The Before-After comparison was based on similar data about a large number of crossings where the warning device has been changed sometime during 1981-1983. We prepared two predictions of 'what would have been without treatment'; one using the Method of Sample Moments, the other using the Multivariate Regression Method. The main results are given in Table 11.9.

Table 11.9. Estimated safety effect.

	Crossbucks to Flashers	Crossbucks to Gates	Flashers to Gates
No. of converted crossings	891	1037	934
No. of 'before' accidents*	165.0	239.1	285.7
Expected 'after' accidents by:			
a. Method of Sample Moments	99.4	150.8	202.1
b. Multivariate Method	100.8	162.0	208.0
Recorded 'after' accidents	49	50	114

* Since the number of 'before' and 'after' years was unequal, the number listed is prorated to the number of 'after' years [i.e., the $\Sigma r_d(j)K(j)$ of Chapter 7].

Using these results, it appears, e.g., that on the sites converted from crossbuck signs to train activated flashers, the reduction from 165 to about 100 was an artifact of regression to the mean. Perhaps the decision to enhance the warning device at some crossings was triggered by the occurrence of an accident there and the very same accidents were included in the 'before' count. Only the reduction from about 100 to 49 is attributable to the conversion. Similar observations pertain to the other two treatments. It also appears, that the two methods for estimating E$\{\kappa\}$ and VAR$\{\kappa\}$ gave similar results for all three treatments.

11.8 NAIVE AND C-G STUDIES REVISITED

The two cases examined in the preceding section are adaptations of the Naive and Comparison-Group study forms discussed in Chapters 7 and 9 of Part II. In the Naive and C-G studies, κ has always been estimated by K and the variance of $\hat{\kappa}$ has also been estimated by K. In contrast, by the EB method, κ is estimated by Equation 11.1 and the variance of $\hat{\kappa}$ by Equation 11.3. The purpose of this section is to present the EB adaptation of the Naive and the Comparison-Group study forms. The adaptation to be added here is in the manner in which κ is estimated.

a. The EB Naive Study

The setting is as follows. Some treatment has been implemented on entities numbered 1, 2,. . ., j, . . ., n. During the 'before' periods the accident counts were K(1), K(2), . . . , K(n) and during the 'after' periods the accident counts were L(1), L(2), . . . , L(n). The duration of the 'before' and 'after' periods may differ from entity to entity. We defined the 'ratio of durations' for entity j to be $r_d(j)$=(duration of 'after' period for entity j)/(duration of 'before' period for entity j).

For each entity there is a reference population[1]. For each reference population we have the estimates $\hat{E}\{\kappa(j)\}$ and $V\hat{A}R\{\kappa(j)\}$, j=1, 2, . . . , n. Using these, we first calculate the 'weights'

$$\alpha(j)=\frac{1}{1+\frac{V\hat{A}R\{\kappa(j)\}}{\hat{E}\{\kappa(j)\}}} \qquad \ldots 11.27$$

and then the estimates

$$\hat{\kappa}(j)=\alpha(j)\hat{E}\{\kappa(j)\}+[1-\alpha(j)]K(j) \qquad \ldots 11.28$$

$$V\hat{A}R\{\hat{\kappa}(j)\}=[1-\alpha(j)]\hat{\kappa}(j) \qquad \ldots 11.29$$

The four steps of statistical analysis are as introduced in Section 6.1. The needed estimates $\hat{\lambda}$, $\hat{\pi}$, $V\hat{A}R\{\hat{\lambda}\}$ and $V\hat{A}R\{\hat{\pi}\}$ are given in Table 11.10.

---

[1] Several treated entities may have the same reference population.

Table 11.10. The estimates for STEPs 1 and 2 of the four-step.

Estimates of Parameters STEP 1	Estimates of Variances STEP 2
$\hat{\lambda}=\Sigma L(j)$	$V\hat{A}R\{\hat{\lambda}\}=\Sigma L(j)$.
$\hat{\pi}=\Sigma r_d(j)\hat{\kappa}(j)$	$V\hat{A}R\{\hat{\pi}\}=\Sigma r_d(j)^2\, V\hat{A}R\{\hat{\kappa}(j)\}$

Using the results in Table 11.10, estimates of $\bar{\delta}$, $\bar{\theta}$, $VAR\{\hat{\delta}\}$ and $VAR\{\hat{\theta}\}$ are calculated, as in Table 9.18, by Equations 6.1 to 6.4 reproduced here. Since 'n' entities are thought off as one 'composite entity', the estimates of $\bar{\delta}$, θ, $VAR\{\hat{\delta}\}$ and $VAR\{\hat{\theta}\}$ pertain to this composite entity.

Table 11.11. STEPS 3 and 4.

$$\bar{\delta} \equiv \pi - \lambda \qquad \text{...6.1}$$

$$VAR\{\hat{\delta}\} = VAR\{\hat{\pi}\} + VAR\{\hat{\lambda}\} \qquad \text{...6.2}$$

$$\bar{\theta}^* = (\lambda/\pi)/[1+VAR\{\hat{\pi}\}/\pi^2] \qquad \text{...6.3}$$

$$VAR\{\hat{\theta}\} \approx \theta^2[(VAR\{\hat{\lambda}\}/\lambda^2)+(VAR\{\hat{\pi}\}/\pi^2)]/[1+VAR\{\hat{\pi}\}/\pi^2]^2 \qquad \text{...6.4}$$

The use of these results is illustrated in Numerical Example 11.7.

Numerical Example 11.7. Conversion from two to four-way stop control in Michigan.

Ten rural intersections in Michigan were converted from two-way to four-way stop control. The 'before' and 'after' periods were of equal duration but not all conversions were in the same years. The count of before accidents is in column 4 of Table 11.12. The total number of 'before' accidents was 146. The count of 'after' accidents is listed in column 5; their total was 33. The reduction from 146 to 33 is almost certain exaggerated by the tendency to select high accident locations for conversion. The Method of Sample Moments has been used to estimate $E\{\kappa\}$ and $VAR\{\kappa\}$. The $\bar{K}$, measured in [accidents/year], and s^2, measured in [accidents/year]^2, are listed in columns 6 and 7. These are based on accidents occurring at all rural two-way stop controlled intersections in Michigan in the corresponding years - the reference population. The entries differ from site to site when the 'before' periods differ. Column 8 gives the number of years in the 'before period'. The entries in columns 9 and 10 are calculated by $\hat{E}\{\kappa\}=r\times\bar{K}$ and $V\hat{A}R\{\kappa\}=r^2\times(s^2-\bar{K})$. Columns 11 to 13 are then computed using Equations 11.27 to 11.29.

Numerical Example 11.7. Conversion from two to four-way stop control in Michigan.

Table 11.12. Data and computations for 10 intersections.

1 Site	2 Before	3 After	4 K	5 L	6 $\bar{K}$	7 s^2	8 r	9 $\hat{E}\{\kappa\}$	10 $V\hat{A}R\{\kappa\}$	11 $\hat{\alpha}$	12 $\hat{\kappa}$	13 $VAR\{\hat{\kappa}\}$
1	71-73	75-77	14	6	0.092	0.151	3	0.28	0.53	0.34	9.30	6.12
2	73-75	77-79	16	3	0.091	0.146	3	0.27	0.5	0.35	10.43	6.73
3	71-73	75-77	18	6	0.092	0.151	3	0.28	0.53	0.34	11.94	7.85
4	71-73	75-77	28	7	0.092	0.151	3	0.28	0.53	0.34	18.51	12.18
5	71-73	75-77	15	3	0.092	0.151	3	0.28	0.53	0.34	9.96	6.55
6	72-74	76-78	28	1	0.091	0.153	3	0.27	0.56	0.33	18.87	12.65
7	75-76	78-79	4	0	0.093	0.145	2	0.19	0.21	0.47	2.19	1.15
8	71-73	75-77	11	3	0.092	0.151	3	0.28	0.53	0.34	7.33	4.82
9	75-76	78-79	6	2	0.093	0.145	2	0.19	0.21	0.47	3.24	1.71
10	72-74	76-78	6	2	0.091	0.153	3	0.27	0.55	0.33	4.11	2.76
Sums			146	33							95.89	62.53

Thus,

STEP 1: $\hat{\lambda}$=33 accidents ; $\hat{\pi}$=95.89 accidents

STEP 2: $V\hat{A}R\{\hat{\lambda}\}$=33 accidents^2; $V\hat{A}R\{\hat{\pi}\}$=62.53 accidents^2

STEP 3: $\hat{\delta}$=95.89-33=62.89 accidents; $\hat{\theta}$ =(33/95.89)/(1+62.53/95.89^2)= 0.34

STEP 4: $V\hat{A}R\{\hat{\delta}\}$=62.53+33=95.53 accidents^2, $\hat{\sigma}\{\hat{\delta}\}$= 9.77 accidents;

$V\hat{A}R\{\hat{\theta}\} \approx 0.34^2\ (33/33^2 + 62.53/95.89^2)/\ (1+62.53/95.89^2)=0.0043,\ \hat{\sigma}\{\hat{\theta}\}=0.066.$

b. The EB Comparison-Group study

The circumstance here is that a treatment has been applied to several entities that may not have the same 'before' and 'after' periods and environment. Therefore, each treated entity (or group of treated entities with common environments and 'before'-'after' periods) has to have a separate comparison group. Thus, once again, we have 'n' treated entities (or groups of entities) labeled 1, 2, . . . , j, . . . , n. For each entity one has to obtain the estimates in Table 11.13. This table is the same as Table 9.17, except that $\hat{\kappa}$(j) replaces K(j), and VAR$\{\hat{\kappa}(j)\}/(\hat{\kappa}(j))^2$ replaces 1/K(j)

Table 11.13. Estimates for STEPs 1 and 2 in an EB Comparison-Group study.

Estimates of Parameters STEP 1	Estimates of Variances STEP 2
$\hat{\lambda}(j)=L(j)$	$V\hat{A}R\{\hat{\lambda}(j)\}=L(j)$
$\hat{r}_T(j)=\hat{r}_C(j)=(N(j)/M(j))/(1+1/M(j))\approx N(j)/M(j)$	$V\hat{A}R\{\hat{r}_T(j)\}/r_T^2(j)\approx 1/M(j)+1/N(j)+V\hat{A}R\{\omega(j)\}$
$\hat{\pi}(j)=\hat{r}_T(j)\hat{\kappa}(j)$	$V\hat{A}R\{\hat{\pi}(j)\}\approx\hat{\pi}^2(j)[VAR\{\hat{\kappa}(j)\}/(\hat{\kappa}(j))^2+V\hat{A}R\{\hat{r}_T(j)\}/r_T^2(j)]$

As in Chapter 6, (Equations 6.5 and 6.6) we define $\lambda \doteq \Sigma\lambda(j)$, $\pi \doteq \Sigma\pi(j)$, $VAR\{\hat{\lambda}\} \doteq \Sigma VAR\{\hat{\lambda}(j)\}$ and $VAR\{\hat{\pi}\} \doteq \Sigma VAR\{\hat{\pi}(j)\}$, where Σ denotes summation over all n entities. Steps 3 and 4 are the usual equations repeated in Table 11.11.

11.9 ADDITIONAL APPLICATIONS

I will show now that the EB method of safety estimation is of interest not only when one inquires about the effect of some treatment but also in other circumstances of practical interest. The setting examined is that of checking whether the κ of an entity is in some sense deviant.

Numerical Example 11.8. Safety of a signalized intersection.

A certain four-legged signalized intersection in Metropolitan Toronto has recorded 5 accidents between 'straight-through' and 'left-turn' vehicles in the last 3 years, on weekdays, between 7 and 9 a.m.. The average traffic flows on weekdays between 7 and 9 a.m. were 450 straight-through vehicles/hour and 120 left-turning vehicles/hour. What is the estimate of the expected frequency of weekday morning-peak accidents between vehicles of these traffic flows for this intersection and how does it compare to what is 'normal'?

These are simple questions. Still, a traffic engineer would find it difficult to give a satisfactory answer. Some might estimate κ to be 5/3 accidents/year. But this disregards all 'clues of the first kind' such as: this is a four-legged signalized intersection, in Toronto, the accidents are between a left-turning and a straight-through vehicle, the corresponding flows are 450 and 120 vehicles per hour. To capture these clues to κ, one has to have an estimate of $E\{\kappa\}$ and $VAR\{\kappa\}$ for a reference population with these traits. However, in all Metropolitan Toronto there are but a few signalized intersections with very similar flows. Thus, no useful real reference population exists. Therefore, the traffic engineer would find it even more difficult to answer the second part of the question, namely, to say what is 'normal' under these conditions. This must be a severe handicap for engineering practice. If it is not known what 'normal' is, how can one judge what is 'deviant'? The Multivariate Regression Method helps to alleviate these difficulties as is shown below.

Answer 1: What is normal.

A multivariate model has been fitted to data from 145 signalized intersections in Metropolitan Toronto (Hauer et al., 1989a). For accidents of this type the model equation is $\hat{E}\{\kappa\}=0.0283\times10^{-6}$(straight-through flow)$\times$(left-turn flow)$^{0.5163}$ accidents/hour. For an intersection with 450 straight-through vph and 120 left-turning vph, $\hat{E}\{\kappa\}=150.8\times10^{-6}$ accidents/hour. With 261 weekdays in a year, in three years there are $2\times261\times3=1566$ hours in the 2-hour morning peak. Thus, in three years, intersections with similar traits should record an average of $150.8\times1566\times10^{-6}=0.236$ morning-peak accidents between straight-through and left-turning vehicles.

This answers the question of what is normal. Although no real reference population existed, it proved possible to say how many accidents of this type should be expected at an average intersection with these specific flows. To do that, a multivariate statistical model had to be fitted to data. Now the count of five morning-peak left-turn accidents in three years can be seen in a new perspective.

However, the question of what is the expected number of such accidents at this specific intersection is still without an answer. The essence of the Multivariate Regression Method now comes into play. It provides a way to estimate VAR{κ}. Note that:

$$\text{If } VAR\{\kappa\}=[E\{\kappa\}]^2/b,$$
$$\text{then } \alpha \doteq (1+VAR\{\kappa\}/E\{\kappa\})^{-1} = (1+E\{\kappa\}/b)^{-1}.$$

Answer 2: The safety of this intersection.

When the multivariate model has been fitted to the Toronto intersections, it has been found that for this type of accident, $\hat{b}=1.39$. Thus, $V\hat{A}R\{\kappa\}=0.236^2/1.39=0.040$ [accidents in three years]^2. As a consequence, the estimate of α is $(1+0.236/1.39)^{-1}=0.855$. Accordingly the estimate of the expected number of such accidents at this intersection is $\hat{\kappa}=0.855\times0.236+0.145\times5=0.202+0.725=0.927$ morning-peak accidents in three years (see Equation 11.1).

This completes the answer to the questions originally posed. The estimate of the expected number of morning-peak accidents between straight-through and left-turning traffic in the 3-year period at this intersection was 0.927, while an average intersection with these flows should be expected to have 0.236 such accidents in the same period.

It is perhaps useful to carry the numerical example a step further and ask whether this difference is a clear sign that our intersection is deviant.

Answer 3: Is this intersection deviant?

An average intersection of this kind is estimated to have 0.236 accidents in three years with a standard deviation of $\sqrt{0.04}=0.2$. If deviant intersections are those with κ, say, two standard deviations above the mean, sites with $\kappa>0.64$ will be judged deviant. The best estimate of the κ for our intersection is 0.927 accidents in three years, which should bring it into the orbit of suspected deviants. However, by Equation 11.3, $\widehat{VAR}\{\hat{\kappa}|5\}= 0.145\times0.927= 0.134$ [accidents in three years]^2 or $\hat{\sigma}\{\hat{\kappa}|5\}=\sqrt{0.134}=0.366$. Thus, there is a good chance that its κ is actually less than 0.64. With a bit of courage and using tables of the standard normal distribution the probability that the κ of the intersection under scrutiny is below 0.64 accidents in three years , can be calculated to be about 20%.

Originally, the EB method of safety estimation has been introduced as a remedy to the regression-to-mean problem. It turns out, however, that its uses are much broader. It is a coherent way to estimate the safety of individual entities whenever the need to do so arises. Important extensions to the uses already shown will be introduced in the next chapter.

11.10 CHAPTER SUMMARY

This chapter is mainly about the estimation of what κ in the 'before' period was. In the early 1980s it became evident that in observational studies one may not assume that the accident count K is a good estimate of κ. After some false starts the Empirical Bayes approach to estimation emerged as the favored tool for avoiding the dreaded regression-to-mean bias. The EB approach is more than a trick to avoid RTM bias; it is a method of estimating κ that coherently exploits not only accidents counts, but also information contained in the traits of an entity. With the right reference population the EB approach gives estimates that are not only free of the RTM bias but are also more precise than what is possible when only accident counts are used.

While the concept of reference population can be clearly defined, whether a population for which data is available is the 'right reference population' is not yet sufficiently clear. Some problems associated with the use of reference populations are discussed in Section 11.3 and the

Endnote. I have argued that the problems arising from the use of a reference population are not substantially different from the problems that arise when one uses accident counts only.

The mathematics of mixing the two kinds of clues (Section 11.4) is not only simple but also logically pleasing. To use the suggested machinery requires that we have estimates of the mean and variance of κ in the reference population. Two methods for the estimation E{κ} and VAR{κ} were offered for consideration (Section 11.5). Of the two, it appears that the Multivariate Regression Method is more practical (in terms of kind of data necessary) and better in use (since it allows that the traits of the treated entity and its reference population be matched exactly). That the EB method is not subject to the RTM bias follows by logic: if an entity registered K accidents we should estimate its κ by the estimate of E{κ|K}. This is what the EB approach provides. That it in fact does the trick is shown by an illustration in Section 11.6. Two cases of application are reviewed in Section 11.7. The main concern in these studies was to cleanse results of the RTM bias. Both make use of a fixed-duration before period and imperfectly account for the effect of the various factors that change with time. A formal EB adaptation of the 'Naive' and the 'Comparison-Group' study forms is provided in Section 11.8. That the EB method has applications beyond the domain of observational before-after studies is indicated in Section 11.9.

In the introduction to Part III I spoke about the conflict between the real circumstances of observational studies and the conventional approaches used for their interpretation. Three principal problems were identified: (a) - the RTM bias that comes from non-random selection for treatment; (b) - the use of a fixed-duration 'before' period which is kept short to make the assumption of unchanging κ plausible; and (c) - the less than satisfactory methods of projecting κ into the future so as to account for changes in traffic flow and other factors.

In the present chapter the use of the EB method has been suggested as a way to avoid the RTM bias. This solves the first problem. However, the EB method has been developed and applied for a fixed-duration before period. Therefore, next task is to rid analysis of this constraint. Doing so will make estimation both better and more logically coherent.

Endnote.

I have said that Smith is male, 22 years old, licensed to drive in Ontario and that, based on the officila statistics, such people had, on the average, 0.1152 accidents in 1984. If that is all that is known about Smith, the best estimate of his safety in 1984 is 0.1152.

To estimate Mr. Smith's safety, it would be important to know whether he was single or married in 1984. Married men tend to have fewer accidents. However, marital status is not in the data base that Ontario maintains. Therefore, this trait is not known for Mr. Smith. Nor is it possible to ascertain from the Ontario data base what is the mean number of accidents for a corresponding reference population. Thus, what has earlier been offered as a legitimate estimate of Smith's safety in 1984, is a reflection of what happens to be in the data base. Had different information been collected one would have made different assertions about Mr. Smith's safety. Thus, **the first difficulty is that the estimate of safety depends on what data happen to be available.**

While marital status is known to be related to a person's safety, there may be other traits that influence safety but remain unrecognized. Thus, the estimate of safety is circumscribed by the extent of the analyst's intuition and understanding. Accordingly, **the second difficulty is that the estimate of safety depends on what is deemed relevant.**

Suppose now that Mr. Smith reveals that in 1984 he was married. We know that 22 years old, male, Ontario drivers, whose marital status is unknown, had on the average 0.1152 accidents in 1984. We also suspect the married and unmarried members of this group have very different averages. Since it is now known that Smith was married, should we still estimate Smith's safety to be 0.1152 accidents/year. If not, what should we estimate his safety to be? Thus, **the third difficulty is that for any specific entity it is always possible to think of it as having some relevant trait which sets it apart from all available reference populations.**

In conclusion, it is sensible to estimate safety by exploiting all useful information, including that which resides in the traits of an entity - the clues of the first kind. However, the use of clues of the first kind is tied to the use of a reference population and this cannot be divorced from the use of judgement. Many will see this as a decisive reason for not using clues of the first kind in estimation. Therefore, to come to a balanced view, the blemishes of clues of the second kind - the accident record - also need to be listed.

The second kind of clue that bears on safety estimation is the historical accident record of the entity. It has been said earlier that Mr. Smith has had no accidents in 1984. Surely this tells something about his safety in 1984. If the count of accidents in 1984 were the only basis for the statistical estimation of safety, we would have to estimate Smith's expected number of accidents in 1984 to be zero. This makes no good sense. Only a person not driving has a zero chance of being in an accident as driver. It is possible to decline making an estimate of safety in these circumstances, insisting that the count of accidents has to be substantial before a statistical estimate can be of any interest. Such a position has the unfortunate consequence that many entities about the safety of which we inquire (drivers, grade crossings, road sections, airports) never have an accident count which can be deemed substantial. If so, one would have to concede that nothing can be said about their safety.

If the option of remaining silent about Smith's safety is unpalatable, it might be possible to make use of Mr. Smith's earlier accident record, hoping that he has had at least one accident in the past. This might avoid the embarrassment of having to estimate that his expected number of accidents in 1984 was zero. Indeed, I said that there was one accident in 1982. Unfortunately, it is not clear how an accident in 1982 is to affect the estimate of Smith's safety in 1984. Surely, Smith's safety changes with the passage of time. Consequently it is necessary to make some assumptions about **how** it changes, before the 1982 accident can be brought to bear on the task of estimating his safety in 1984. The assumption made most frequently and tacitly is that his safety did not change; it has the merit of being convenient but lacks the virtue of being sensible. I will return to this issue later.

If one wishes to use Smith's earlier accident record to estimate his safety in 1984, judgement must enter into estimation in much the same manner as when clues of the first kind are used. To see why, assume that Smith's mileage in 1982 and 1984 has something to do with his chance of being in an accident and this should be taken into account when using 1982 data to estimate his safety in 1984. However, data about Smith's mileage are not kept. Thus, just as before, the first problem is that **the estimate depends on what data happen to be available.**

Second, while exposure has been recognized as related to Smith's safety there must be other conditions of Smith's life which are also related to his safety and which were not the same in 1982 and 1984. It is not that these have not been measured, we just do not know what these conditions are. Thus, **the second difficulty is that the estimate of safety depends on what is thought to be relevant**.

Suppose now that it is possible to account for the change in some common 'conditions' such as, the amount of driving. However, Smith tells that he graduated from school in 1982 and was a salesman in 1984. How this affects a person's safety is not known. Thus, just as before, the third difficulty is that for any specific entity **it is always possible to think of some change in conditions that are peculiar to this entity and may affect its safety in an unknown manner.**

These difficulties are just the same as were encountered when safety estimation was based on traits and a reference population. They are rooted in the inconvenient fact that the safety of each entity changes in time. Therefore when safety estimation is based solely on the accident history of an entity, estimates depend on the arbitrary and the judgmental in precisely the same manner as when estimation is based on traits and reference populations. To estimate Smith's safety only from his accident record, as if we did not know that he has been driving in Ontario, was male and was 22 years old, is to disregard important information. Conversely, to consider Smith's traits but not to take note of his actual accident record also amounts to discarding relevant information. It ought to be obvious that it is best to use both kinds of clues: those which derive from traits and also those which derive from the count of accidents.

At the same time, it is necessary to recognize that if the traits of entities are to affect estimation, or if the safety of an entity changes in time, the process of safety estimation will be unlike that of measuring distance, pressure, or rate of reaction. Safety estimation will depend on what data are available, on what is recognized as relevant, and on other manifestations of judgment. One should not pretend otherwise.

CHAPTER 12

A MORE COHERENT APPROACH?

The programme for Part III of this monograph is to suggest improved methods for interpreting observational Before-After studies. In Chapter 11 I have indicated how the Empirical Bayes approach can help to obviate the regression-to-mean (RTM) bias that comes from the non-random selection of entities for treatment. But this is only a part of a larger task. The overall aim is to put the prediction of 'what would have been the safety in the after period had the treatment not been implemented' on firmer footing. To do that, I have to do away with several additional impediments inherited from various traditions:

Impediments:

a. The notion that there is a fixed-duration 'before' period and that accident counts in years prior to the onset of this 'before' period contain no useful information;

b. The convention that accident counts from such a fixed-duration 'before' period are all to be used to estimate a single expected value (κ), and the belief that this single value can then be 'suitably modified' to serve as a prediction for the 'after' period;

c. The commonly held belief that the best way to 'suitably modify' the estimate of κ is by using a comparison group.

To set out in the right direction, imagine that for an entity there is a sequence of values κ_1, $\kappa_2, \kappa_3, \ldots, \kappa_Y$ for years 1, 2, 3, . . . , Y where Y is the last year before the treatment. These κ's are the foundation for the prediction of 'what the κ's of this entity would have been' in the after-treatment period, namely, the $\kappa_{Y+1}, \kappa_{Y+2}, \ldots, \kappa_{Y+Z}$. In this, Z is the number of years after treatment for which we wish to predict. It must seem obvious that to predict well one should use an estimate of a fairly long time series $\kappa_1, \kappa_2, \kappa_3, \ldots, \kappa_Y$, not merely the average of the last two or three terms. These items form the lower tier of Figure 12.1.

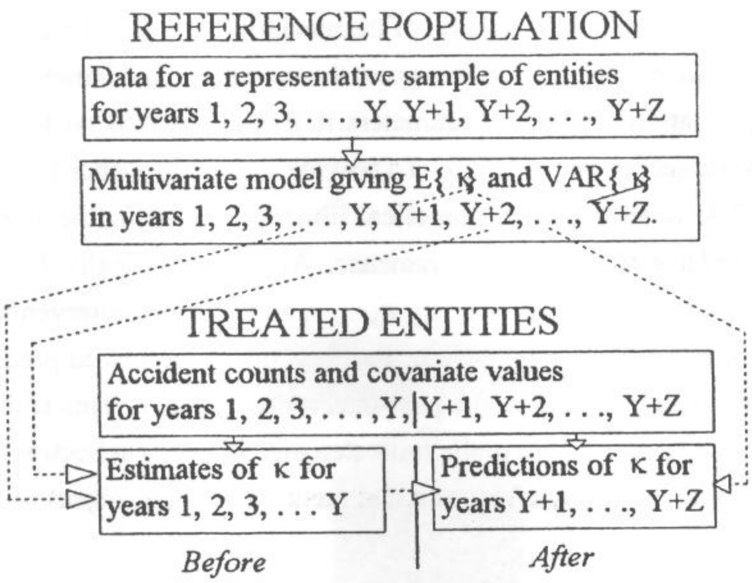


Figure 12.1. Main elements of approach.

The procedure will be based on the EB approach. After some modification, the concepts already introduced in Chapter 11 will prove useful in this, more general, setting. As always, the key problem is to predict what would have been the $\kappa_{Y+1}, \kappa_{Y+2}, \ldots, \kappa_{Y+Z}$ in the 'after' period for each treated entity, had the treatment not been applied. The EB approach is rooted in the use of a multivariate model for a reference population[1] to fashion these predictions. Thus, the multivariate model for entities of the reference population, shown as the upper tier in Figure 12.1, will serve several purposes. Note that it spans the 'before' and 'after' periods. One purpose, as in Chapter 11, is to yield estimates of $\kappa_1, \kappa_2, \kappa_3, \ldots \kappa_Y$ for each treated entity, so that the estimates are free of the regression-to-mean bias and are statistically efficient. These estimates are the launching pad for the leap into the unobserved future - the prediction. The change in the κ's from an observed 'before' period to an imagined 'after-period-with-no-treatment' is due to change in two kinds of factors: factors that are measured and the influence of which can be accounted for explicitly; and factors that are unmeasured or the influence of which is ill-understood. Change due to both kinds of factors can be represented in a multivariate model. Thus, the multivariate model will be used to account for the change in both types of factors on each treated entity. In this manner, the multivariate model for a reference population will serve both EB estimation (in years 1, 2, . . . ,Y) and prediction (for years Y+1, Y+2, . . . , Y+Z).

---

[1] There may be other attractive options, but these will not be pursued here.

I will elaborate on the uses of multivariate models of accident counts in Section 12.1. Section 12.2 will be devoted to the elucidation of the meaning, form and the assumption implicit in the model equation and its parameters. At this point the mathematical element of the plot begins to thicken. For the estimation of model parameters it is necessary to write down a 'likelihood function'. This is the subject of Section 12.3. To provide a tangible anchor for the tangled algebra, the next Section (12.4) is an extended numerical illustration of how the likelihood function is computed and then used to estimate model parameters. At this point, finally, I can reach for the next rung on the ladder. I will show how the $\kappa_1, \kappa_2, \kappa_3, \ldots, \kappa_Y$ of the pre-intervention period are to be estimated. This is the subject of Section 12.5. The next question, how to predict $\kappa_{Y+1}, \kappa_{Y+2}, \ldots, \kappa_{Y+Z}$ or 'what the κ would have been' in the post-intervention period. This is discussed in Section 12.6. This will be followed by a return to the four- step of Chapter 6 (in Section 12.4). How all this has been used to estimate the safety effect of a treatment will be the subject of Section 12.7.

12.1 USES OF MULTIVARIATE MODELS OF ACCIDENT COUNTS

Multivariate models for accident counts have many uses. Here we are concerned only about their use in observational Before-After studies. The sections that follow contain some conceptual innovation and tedious mathematics. So as not to lose sight of the main purpose in a maze of detail, I will restate it briefly at the outset. A treatment has been applied to 'n' entities. For each of the treated entities we are to predict what would have been the expected accident frequencies in the 'after' period. These 'n' predictions are then compared to 'n' estimates of what the accident frequencies in the 'after' period were. Conclusions about the effect of the treatment on safety are based on this comparison. The main challenge is to prepare the required predictions. At the beginning of this chapter I listed three shortcomings of the way predictions are now being made. Multivariate modeling of data from reference populations will help remove these shortcomings in three ways.

To begin with, multivariate models of data from reference populations are the source of estimates of E(κ} and VAR{κ} that are needed for Empirical Bayes (EB) estimation (see the upper tier in Figure 12.1). It is EB estimation that helps to cleanse the final result from the regression-to-mean bias. Specifically, consider a treated entity 'i' (lower tier in Figure 12.1) thought to have in the 'before' years 1, 2, 3, ..., y, ..., Y expected accident frequencies $\kappa_{i,1}, \kappa_{i,2}, \kappa_{i,3}, \ldots, \kappa_{i,y}, \ldots, \kappa_{i,Y}$. In this, 'Y' is the last year before treatment. The task is to **prepare estimates** of the $\kappa_{i,1}, \kappa_{i,2}, \kappa_{i,3}, \ldots, \kappa_{i,y}, \ldots, \kappa_{i,Y}$, and to later **use them to predict** what $\kappa_{i,Y+1}, \kappa_{i,Y+2}, \ldots, \kappa_{Y+Z}$ would have been in the 'after treatment' years y=Y+1,Y+2,..., Y+Z, had the treatment not been implemented. To estimate the $\kappa_{i,1}, \kappa_{i,2}, \kappa_{i,3}, \ldots, \kappa_{i,y}, \ldots, \kappa_{i,Y}$ by the EB method, we need not only the accident counts $K_{i,1}$, $K_{i,2}$, $K_{i,3}$, . . . , $K_{i,y}$, . . . , $K_{i,Y}$ but also estimates of the means $E\{\kappa_{i,y}\}$ and the variances $VAR\{\kappa_{i,y}\}$ for the reference population of entity 'i'. That estimates of $E\{\kappa_{i,y}\}$ and $VAR\{\kappa_{i,y}\}$ can be obtained from a multivariate statistical model[1] has already been shown in Section 11.5. In this section I will exploit the foundations already laid and will put more flesh on the bone.

Second, multivariate models for reference populations will provide the means to overcome another thorny problem. To illustrate the problem, recall that in Section 11.3 I spoke of a certain Ontario driver, Mr. Smith, whose κ in 1984 ($\kappa_{Smith,\ 1984}$) was to be estimated. Mr. Smith was said to have had an accident in 1982. I noted there, that one may not assume that his κ in 1982 was the same as in 1984. Therefore, I said, it is not clear without further elaboration how the evidence of his accident in 1982 is to influence the estimate of $\kappa_{Smith,\ 1984}$. It is as if a two-year old record of Smith's blood pressure was to be used to assess his blood pressure now. While the two are linked, pertaining to the same person, one may not take them to be from the same distribution.

This, is a general problem. Much of the theory of probability and statistics is about how to use outcomes of repeated trials for statistical estimation. But such estimation is possible only as long as the relevant conditions remain fixed from trial to trial. In our case, the 'conditions of the trial' are changing since one must assume that the expected number of accident counts for any entity is changing from year to year. If each year has its own 'conditions', then each year is a new 'trial' and these trials cannot be repeated. The question arises: how can one use the information contained in the outcome of one trial (the accident count in one year) to make inferences about the expected value (the expected accident count, κ) in another year? In other words, the question is how to use the evidence from a time series[2] of annual accident counts to estimate the expected number of accidents in a certain year, given that the expected values are changing from year to year. Only if this difficulty can be resolved, will it prove possible to abandon the use of short 'before' periods; only then will one be in a position to use a longer time series of accident counts and to build a solid base for the 'prediction'.

---

[1] It is quite common to estimate the safety of entities by multivariate statistical models. Thus, e.g., there are multivariate models to estimate the safety of a grade crossing as a function of the train and car traffic, location, geometric features, and the type of warning device used. Similarly, there are models that estimate the safety of a driver as a function of age, gender and number of convictions for offenses associated with the use of the road. Multivariate models exist also for estimating the safety of intersections and roads. See also Section 11.5.

[2] As the text suggests, this kind of problem is the subject of a statistical specialty area known as 'Time Series Analysis'.

Specifically, we wish to make use of the entire time series of accidents counts $K_{i,1}$, $K_{i,2}$, $K_{i,3}$, . . . , $K_{i,y}$, . . . , $K_{i,Y}$ to estimate any member of the time series $\kappa_{i,1}$, $\kappa_{i,2}$, $\kappa_{i,3}$, . . . , $\kappa_{i,y}$, . . . , $\kappa_{i,Y}$. To be able to do so, we need to understand how the κ's change as traffic and other factors change over the years. This understanding too can be perhaps obtained from a multivariate statistical model.

Finally, the third use of a multivariate model for data from a reference population is in prediction. That is, the estimates of the 'before' time series $\kappa_{i,1}$, $\kappa_{i,2}$, $\kappa_{i,3}$, . . . , $\kappa_{i,y}$, . . . , $\kappa_{i,Y}$ serve as a launching pad for predicting the 'after' time series $\kappa_{i,Y+1}$, $\kappa_{i,Y+2}$, . . ., κ_{Y+Z} . In this case the multivariate model and the reference population which it represents, play a role similar to that of a 'comparison group'. It will help to account for the effect of change from 'before' to 'after' in all causal factors.

Thus, I propose to develop a multivariate model using data for untreated entities in the Y+Z years. The results of this modeling effort will feature in three stages of the suggested approach to the interpretation of observational Before-After studies:

a. For a treated entity 'i', the estimation of $E\{\kappa_{i,y}\}$ and $VAR\{\kappa_{i,y}\}$ of its reference population;

b. For a treated entity 'i', the use of its time series of accident counts $K_{i,1}$, $K_{i,2}$, . . . , $K_{i,Y}$ to estimate the time series $\kappa_{i,1}$, $\kappa_{i,2}$, $\kappa_{i,3}$, . . . , $\kappa_{i,y}$, . . . , $\kappa_{i,Y}$ when these expected values are not all the same.

c. For a treated entity 'i', to predict the $\kappa_{i,Y+1}$, $\kappa_{i,Y+2}$, . . . , $\kappa_{i,Y+Z}$.

Because of this central role which the multivariate models play, it is important to dwell on the assumptions behind the multivariate statistical modeling of accident counts and the meaning of the results.

12.2 THE MODEL EQUATION: MEANING, FORM AND ASSUMPTIONS

I will call a 'multivariate statistical model' an equation, or a set of equations, that link the expected accident frequency[1] of an entity to its observed traits. The observed traits or 'covariates' are items such as 'traffic flow', 'road section length', 'intersection density', 'number of lanes, 'Shoulder Width' and the like. There exists a large body of literature on multivariate statistical

[1] Unless stated otherwise I will stick with what is the common measure of accident frequency, namely: number of accidents/year.

modeling in general; how to choose a functional form, how to estimate parameters, how to identify outliers, and so on. There is also specific guidance on the multivariate modeling of accident counts [see, e.g., Maycock and Hall, 1984, Pickering et al., 1986, Miaou and Lum (1993), Maycock and Summersgill (1994), Maher and Summersgill (1996), Kulmala (1995), Mountain and Fawaz (1996)]. I will not attempt to present here an abridgment of what is done better elsewhere. My intent is to discuss only a specific Empirical Bayes point of view and to examine those issues that are specific to the modeling of expected accident frequency.

The circumstance I have in mind is when there exists a data-set for many entities (say, road sections or intersections). The entities for which we have data are taken to be a sample that is representative of a larger population of entities. For each entity in the data set we have accident counts and covariate values for several years. Since for each entity some covariate values change from year to year, the expected accident frequencies of that entity (the κ's) must also be assumed to change from year to year. Emphasis will be on how such a multivariate model fits into the conceptual framework of the EB method (Chapter 11), and how it is linked to the notion of 'imagined reference populations' first introduced in Section 11.5. I will dwell on two of the steps of multivariate statistical modeling:

a. The choice of model form (model equation);
b. The estimation of parameters.

In the course of developing a multivariate model, steps 'a' and 'b' may have to be repeated several times, the result of one cycle of exploration motivating modifications for the next. Once parameters have been estimated and residuals examined, the form of the model may have to be revised, covariates may need to be added or dropped, and parameters estimated anew. Still, steps 'a' and 'b' are essentially sequential; the choice of the model form determines what parameters have to be estimated and therefore in each cycle 'a' precedes 'b'. In the present section I will discuss some important questions which pertain to the choice of model equation. The main issues of parameter estimation will be discussed in the next section.

The core of all multivariate statistical modeling is the belief that the expected accident frequencies (the $\kappa_{i,y}$) are associated in some orderly fashion with 'causal factors'. Section length, traffic flow, lane width, road grade, etc. are commonly used causal factors. This belief is embodied in a **model equation**. A model equation states in what way $\kappa_{i,y}$ is a function of the various covariates that represent the causal factors. Thus, e.g., if section length (d) and traffic flow (F) are the only traits to be represented, for road sections, model equations such as $\kappa=d\alpha F^{\beta}$, $\kappa=d(\alpha F+\beta F^2)$, $\log(\kappa)=\log(d)+\log(\alpha)+\beta\log(F)+[\gamma\log(F)]^2$, $\kappa=d\alpha F^{\text{integer}}e^{\beta F}$ and the like can be useful. In these, the expected accident frequency for the road section κ is the **dependent variable**, the section length 'd' and traffic flow 'F' are **covariates**, and α, β, γ and 'integer' are **parameters**[1]. Model equations often contain many covariates. Thus, e.g., one of the well known model equations for two lane rural roads (based on the work by Zegeer et al., 1987) is: Accidents/year=section length$\times$0.0031(average daily traffic)$^{0.9425}$ $\times 0.897^{\text{lane width}} \times 0.9157^{\text{average paved shoulder width}} \times 0.94^{\text{average unpaved shoulder width}} \times 0.9739^{\text{median recovery distance from edge of shoulder}}$. In this equation the covariates are: section length, average daily traffic, lane width, average paved shoulder width, average unpaved shoulder width and median recovery distance from edge of shoulder[2].

If one can ascribe relative importance to steps 'a' and 'b' of multivariate modeling, the choice of the model equation (step 'a') is more influential in determining the quality of the product than is the statistical technique used to estimate the parameter values (step 'b'). The choice of model equation should reflect prior knowledge and beliefs about the nature of the relationship, and also insights obtained from the analysis of the specific data set on hand. Because of the primacy of this step in multivariate statistical modeling, it is important to dwell on what the choice of a specific form of the model equation implies. It is also important to examine how such implications square with what we believe to be true about relationship between κ, the covariates, and the parameters. To provide a tangible carrier for this discussion - a strawman to be discarded later - consider the following simple model equation:

$$\kappa_{i,y} = d_i \alpha_y F_{i,y}^{\beta} \qquad \ldots 12.1$$

In this equation, $\kappa_{i,y}$ is the expected accident frequency of road section 'i' in year y; d_i is the length of this road section; α_y is a parameter for year y; $F_{i,y}$ is the traffic flow for road section 'i' and year y; and β is a parameter determining how $\kappa_{i,y}$ changes with $F_{i,y}$. Although this model equation is particularly simple, the implications of choosing it are many.

Consider first the road section length (d_i) featuring in Model Equation 12.1 as a constant that multiplies the rest of the expression. This choice embodies the belief that a road section along which

---

1 The parameter 'integer' is chosen to be 1 or 2 to represent a linear or a quadratic dependence of κ with F when F is small. The linear form is thought to suit single-vehicle accidents while the quadratic form may be best for two-vehicle collisions.

2 For intersections, model equations of the form $\alpha(F_1+F_2)^{\beta}$, $\alpha(F_1F_2)^{\beta}$, $\alpha F_1^{\beta}F_2^{\gamma}$, $\alpha(F_1+F_2)^{\beta}\times[F_2/(F_1+F_2)]^{\gamma}$, with F_1 and F_2 the major and minor inflows, are often used (see, e.g., Mountain and Fawaz, 1996).

all traits (traffic, alignment, pavement friction etc.) remain constant, will have a constant expected accident frequency per unit length for all its possible subdivisions[1].

Next, consider the parameters α_y (there will be one for each year). In Model Equation 12.1 the purpose of the α_y's is to capture the influence of all factors that change from year to year, except for the change in traffic flow. The influence of changes in traffic flow is accounted for separately through F^{β}. Thus, the α_y's can be used to account for the joint influences of year-to-year changes in weather, in economic conditions, in the threshold that makes property damage accidents reportable, in the inclination to report accidents, etc. Note that the α_y's change with 'y' (the index of years) but not with the subscript 'i' which identifies a specific road section. The absence of a subscript 'i' reflects two separate beliefs. First, that weather, economic conditions and similar factors changed in the same manner on all road sections in the data set and, by implication, for the entire population which the data set is thought to represent. Second, that the effect of a specific change from year-to-year (say, in weather) affects all road sections in the same manner. That is, the change affects $\kappa_{i,y}$ in the same proportion as it affects $\kappa_{j,y}$. There is yet another assumption behind the choice of the α_y's. As the model is written, we need to estimate parameters $\alpha_1, \alpha_2, \ldots, \alpha_Y, \alpha_{Y+1}, \alpha_{Y+2}, \ldots \alpha_{Y+Z}$. That each year has its own α must reflect a belief that the change from year to year in the factors that the α_y's are to represent, is not smooth. Otherwise it might have been more parsimonious to represent the effect of these factors by some convenient and smooth mathematical function of time that can be described by fewer parameters than the Y+Z values α_y. Thus, the humble multiplicative parameter α_y carries with it a whole bundle of assumptions and implications: that the factors it represents changed in the same manner for all road sections, that all road sections are affected by this change in the same proportion, and that the change of these factors with time is not necessarily smooth.

The inclusion of $F_{i,y}$ (the traffic flow for road section 'i' and year y) in the model also reflects a belief; namely, that differences in traffic flow make for differences in $\kappa_{i,y}$. The exponent β determines the manner in which $\kappa_{i,y}$ is thought to depend on $F_{i,y}$. Since in Equation 12.1 there is only one β, the belief must be that the dependence on traffic flow is the same for all road sections and for all years. The form $F_{i,y}^{\beta}$ chosen here also has important consequences. It guarantees that as $F_{i,y} \to 0$,

---

[1] Some research indicates that κ may not be proportional to d. On examination we found this to be true only when intersection accidents are included in the count of accidents for a road section. Naturally, intersection accidents have to be modeled separately. When intersection accidents were separated out, κ was found to be proportional to d. However, it is also true, that the length of a road section in many data-sets is decided by considerations of homogeneity of geometric or traffic features. Therefore, there may be some interaction between section length and other covariates. In modeling, this ought to be attributed to the geometric and traffic covariates, not to section length.

$\kappa_{i,y} \to 0$. This is as should be, since without traffic there can be no motor vehicle accidents. However, it should also be noted that this form cannot possibly represent a relationship in which, say, $\kappa_{i,y}$ initially increases with traffic flow but, at some point, as traffic flow increases further, accident frequency begins to diminish[1] - the function $F_{i,y}{}^{\beta}$ has no maximum. This is a severe limitation. It shows how the chosen functional form may be an artificial constraint. Instead of being a device for revealing the relationship behind noisy data, an ill-chosen functional for can produce parameters that obscure or misinterpret the underlying regularity. Taken together, these points demonstrate that a specific model equation always reflects a series of assumptions, beliefs and limitations.

The data used in modeling serve as a reality check on the chosen model equation. If a covariate does not improve the fit of the model to the data, its inclusion in the model equation can be reconsidered. Conversely, if the introduction of a new variable improves the fit, perhaps it should be included in the model. Unfortunately, guidance of this nature is usually weak and often ambiguous because the covariates available to us are often causally intervowen and therefore statistically dependent.

Another important aspect to note is the metamorphosis of the model equation that occurs in the course of the modeling activity. When the model equation is first conceived, the modeler aims to capture plausible cause-effect relationships. However, once written down and used to estimate unknown parameters, the 'fitted model equation' cannot be said to capture cause and effect. It can only represent what statistical associations are present in a set of data. In concrete terms, consider a data set in which some road sections have traffic flows around A and, on the average, have $\alpha_y A^{\beta}$ accidents/km-year; while other road sections have traffic flows around B and, on the average, have $\alpha_y B^{\beta}$ accidents/km-year. This is merely an 'association', a statement of what is true about two different sets of road sections, those with traffic flows around A and those with traffic flows around B. This association gives little assurance that if the flow on some specific road section changes from A in year 1 to B in year 2, then its κ is expected to change by the ratio $(\alpha_2 B^{\beta})/(\alpha_1 A^{\beta})$. The reason is, that roads with differing flows are often built to differing standards, serve different trip purposes, vehicles and drivers, differ in maintenance, enforcement etc. If a model was fitted to this data, the model equation $\alpha_y F^{\beta}$ would prove to be a good choice. If one then used F=A and F=B in this model equation, it would seem that it is the change in traffic flow that causes accident frequency to change. But we know that at the level of the original data, the corresponding change in accident frequency may have been also due to a host of factors that are linked to traffic flow. Model equations are commonly used in this questionable cause-effect mode. Grudgingly, I will do the same, later.

---

1 There is evidence to suggest, e.g., that as traffic flow increases beyond a certain value the frequency of single vehicle accidents begins to diminish.

There is yet another assumption hidden in the way the model equation has been written. It is that entities that have the same values for the covariates have the same κ. Thus, were model Equation 12.1 used, the assumption would be that all road sections with the same traffic flow also have the same expected accident frequency per unit length of road. This is surely a gross over-simplification. Many features other than traffic flow distinguish one road section from another. While the covariates included in a specific model may be the main determinants of κ, one cannot hope to include them all. This is why entities that are identical in all covariates will still **not** have the same κ. To have a realistic understanding of what multivariate modeling in this setting is, one has to look more closely at the left-hand-side of the model equation[1].

So far I have discussed the many assumptions lurking behind the right-hand-side of a model equation such as 12.1. A revision is needed of the left-hand-side because it represents and conceals an unrealistic assumption. To show what it is, and to provide a realistic point of departure, imagine a large number of entities, all with the same covariate values in the model equation as those of entity 'i'. As in Chapter 11, this imagined group of entities will be called the **imagined reference population** for entity 'i'. Although all the entities in this group have the same covariates values in the model equation, they still differ in many other respects that are not captured by the model equation. Road sections may differ in roadside development, distance from the city, age of the pavement, speed distribution, road user demography, etc. These are but some covariates that are not usually incorporated in the model equation. Therefore, each entity in this imagined reference population still has its own distinct κ that is different from the κ's of other entities in that imagined reference population. But the way Equation 12.1 has been written, it would seem that all entities with the same model covariates are expected to have the same κ. This is an impoverished view of reality. To properly represent the real state of affairs, I assume from here on that entities with the same model covariate values each have their own κ. Let

$E\{\kappa_{i,y}\}$ denote the mean of the κ's in year 'y' for all entities in the imagined reference population of entity 'i' (that is, in an imagined population of entities that have exactly the same model covariate values as entity 'i') and,

$VAR\{\kappa_{i,y}\}$ denote the variance of these κ's.

We now reorient our view of multivariate statistical modeling as follows. The task is to build a multivariate statistical model using data. The entities for which we have data are considered to be a representative sample from what we imagine to be a very large population of entities. For this sample of entities we have information about accident counts and about various covariates. We

---

[1] See also Section 11.5 b.

intend to use these data to estimate the parameters of a model equation. The product of this effort will be a 'fitted model' which, for chosen values of the covariates, yields estimates of $E\{\kappa_{i,y}\}$, not of $\kappa_{i,y}$. That is, the **fitted model gives estimates of the *mean* for the imagined reference population** of entity 'i'. The $\kappa_{i,y}$ of the specific entity that happens to be in our data set is, in general, different from the $E\{\kappa_{i,y}\}$ that is being estimated by the multivariate model[1]. Thus, e.g., if a model of the kind given by Equation 12.1 was considered appropriate, one should rewrite the model equation as

$$E\{\kappa_{i,y}\} = d_i \alpha_y F_{i,y}^{\beta} \qquad \ldots 12.2$$

While in notation the difference between Equations 12.1 and 12.2 may seem innocuous, consisting of replacing $\kappa_{i,y}$ by $E\{\kappa_{i,y}\}$, this slight change amounts to a shift of paradigm toward an Empirical Bayes point of view. A one-dimensional depiction of the situation is in Figure 12.2 which is the same as Figure 11.3 used earlier.

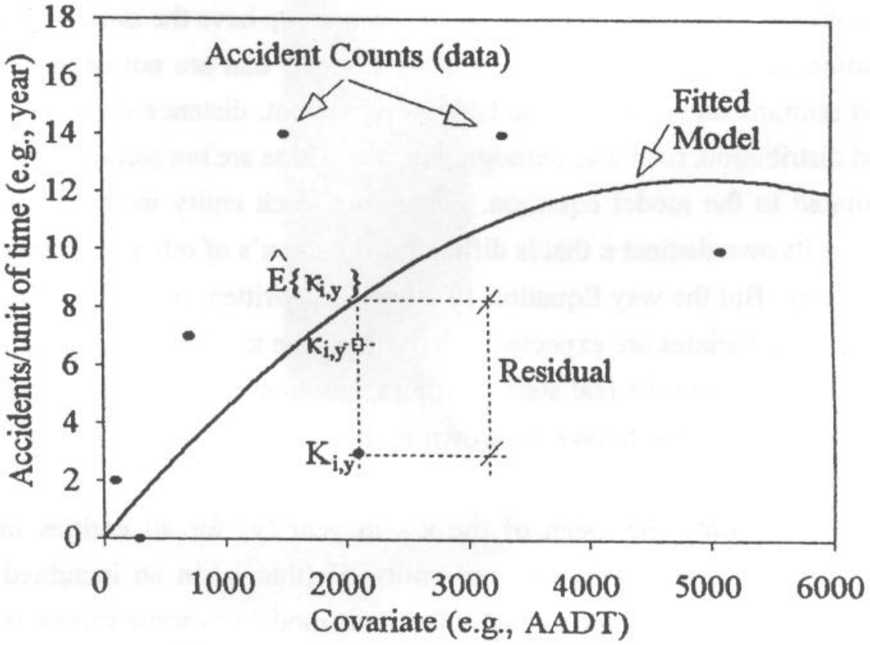


Figure 12.2. Fitted values and residuals.

As noted in Section 11.5b, a residual is the difference between the accident count for year 'y' and entity 'i' ($K_{i,y}$) which served as datum for parameter estimation, and the estimate of the mean

---

[1] Still, if we know nothing about the accident record of road section 'i', the estimate of $E\{\kappa_{i,y}\}$ can also be used to estimate the $\kappa_{i,y}$ of road section 'i' in year 'y'. This is so, because road section 'i' has the same covariates as all other road sections in the imagined group of road sections that has $E\{\kappa_{i,y}\}$ as mean. This is the same argument as the one used for "Smith, the Ontario driver" in Section 11.2.

for the imagined reference population ($\hat{E}\{\kappa_{i,y}\}$) calculated from the fitted model equation. The residual can be thought to consist of several parts. One part is the difference between the expected accident frequency of entity 'i' and the count of accidents on it ($\kappa_{i,y}$-$K_{i,y}$). Another part is the difference between the expected accident frequency of entity 'i' and the estimate of the mean for its reference population ($\hat{E}\{\kappa_{i,y}\}$-$\kappa_{i,y}$). The mean value of the second part, when squared, can serve to estimate VAR$\{\kappa_{i,y}\}$. Thus, viewing the multivariate model as estimating E$\{\kappa_{i,y}\}$, not $\kappa_{i,y}$, allows one to both recognize the existence of VAR$\{\kappa_{i,y}\}$ and also to estimate its magnitude[1].

In summary, the model equation has covariates and parameters on the right-hand side and E$\{\kappa_{i,y}\}$ on the left-hand side. The choice of the model equation reflects beliefs, embodies assumptions, and imposes limitations. Once the model equation is stated and the value of covariates ascertained, the unknown values of parameters can be estimated from data. This task is considered next.

12.3 LIKELIHOOD FUNCTION FOR PARAMETER ESTIMATION

There are two kinds of elements on the right-hand side of a model equation: covariates and parameters. The covariates represent values that describe the traits of an entity such as section length, traffic flow, number of lanes or grade. Thus, e.g., in Equation 12.2, d_i and $F_{i,y}$ are covariates - the length of road section 'i' and the flow of traffic on this road section in year 'y'. The α_y and β in Equation 12.2 are parameters. Parameters are values to be estimated using data. Usually the values of the parameters are so estimated that the model equation fits the data as well as possible.

In what follows, parameter estimation will be by the method of maximum likelihood. That is, an attempt is made to express the probability to observe the accident counts in the data as a function of only the covariates and the unknown parameter values. When the covariate values are then 'plugged into' this expression, the probability to observe the accidents counts in the data is a function of only the unknown parameters. This is the called the 'likelihood function' and will be denoted by $\mathcal{L}$. The central idea is to find those parameter values that make $\mathcal{L}$ largest. That is, of all the possible parameters values, the chosen set of parameters makes the probability to observe the accident counts in the data largest. Accordingly, the main task in this section is to formulate a sensible likelihood function. I will show only the main building blocks from which $\mathcal{L}$ is built and the final result. Details and proofs are relegated to 'Derivations' at the end of the section.

---

1 Actually there is a third part contributed by the difference between E$\{\kappa_{i,y}\}$ and its estimated value. I will disregard this part here but recognize that it will require more careful examination.

The pain of developing the likelihood function from scratch cannot be justified by the need to estimate the parameters of a model equation alone. Most statistical software packages (GLIM, GENSTAT, SPSS, SAS, LIMPED and others) allow the user to obtain maximum likelihood parameters estimates for many model equations[1]. There are, however, three reasons for undertaking this task. First, the formulation here contains the specification of an assumption - a rule linking the year-to-year change in the κ's of an entity to the change in its covariates. This modifies the likelihood function used in the aforementioned software packages and, in general, would lead to somewhat different estimates[2]. Second, in Section 12.5 I will have to use the likelihood function to estimate these κ's. Third, the development of the likelihood function allows me to point out its Empirical Bayes content and the assumptions that go into it.

To begin, I will use the time-honored assumption that the count of accidents on any entity obeys the Poisson probability law. To be specific, consider entity 'i' in year 'y'. The assumption is that

$$P(Accident\,Count = K_{i,y} \mid \kappa_{i,y}) = \kappa_{i,y}^{K_{i,y}} e^{-\kappa_{i,y}} / K_{i,y}! \qquad \ldots 12.3$$

That is, if $\kappa_{i,y}$ is the expected number of accidents per year for entity 'i' in year 'y', then the probability that the count of accidents on this entity and year will be $K_{i,y}$ is given by the expression on the right-hand side of Equation 12.3. From the reference population we have a representative sample of R entities, and for each entity we have information about accident counts for Y years before treatment has been applied to a different set of entities (see Figure 12.1) and for Z years after the treatment has been applied. With this, the probability that the set of accident counts $\{K_{i,y}\}$ on entities i=1, 2, . . . , R and in years y=1, 2, . . . , Y+Z will be observed can be written as the double product:

$$P(Accident\,Counts\,\{K_{i,y}\} \mid parameters\,\{\kappa_{i,y}\}) = \prod_{i=1}^{R} \prod_{y=1}^{Y+Z} \kappa_{i,y}^{K_{i,y}} e^{-\kappa_{i,y}} / K_{i,y}! \qquad \ldots 12.4$$

---

1 The software be can used only for 'generalized linear model equations'. This is that family of equations for which there exists some monotone function 'f' such that $f(E\{\kappa_{i,y}\})$ is a linear function of the covariates. Thus, e.g., Equation 12.2 could be written as $E\{\kappa_{i,y}\}=d_i \exp[\ln(\alpha_y)+\beta\ln(F_{i,y})]$. Using the 'ln' function for 'f', $\ln(E\{\kappa_{i,y}\})= \ln(d_i)+\ln\alpha_y+\beta\ln(F_{i,y})$ which is a linear function of the covariate $\ln(F_{i,y})$.

2 The issue has been discussed in Maher and Summersgill (1996, Section 7). They suggest to use an iterative approach based on a Taylor series expansion. Our approach is to use what seems a sensible assumption (as stated by Equation 12.5) and to write the corresponding likelihood function.

This function contains R×(Y+Z) unknown parameters in the set $\{\kappa_{i,y}\}$. The challenge is to modify Equation 12.4 so as to replace the many unknown parameters $\{\kappa_{i,y}\}$ by the known covariates and the few unknown parameters of the model equation. To illustrate, if we had data for four years on each of six road sections, and if model Equation 12.2 was to be used, the task would be to replace the 24 unknown values of the $\kappa_{i,y}$ by the parameters $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta\}$ and one additional parameter 'b' to be defined shortly[1]. To accomplish this, I have to add two assumptions to the one already made (namely, that accident counts are Poisson distributed).

Embedded in the formulation so far is the notion that for each entity the κ's change from one year to another. This is why instead of a single κ_i for entity 'i' we have several $\kappa_{i,y}$, one for every year. However, nothing has been said so far about **how** the κ's of an entity change from year to year. The next step is to make some sensible assumptions in this matter. Consider a road section which from year to year retains some of its basic features (such as length, number of lanes, alignment or grade). However, some determinants of the safety of this road section (such as traffic, weather, vehicle mix) do change from one year to another. Therefore, one can expect that for this road section there will be some basic similarity between its κ's over the years, but that there will also be some change of the κ's from year to year. Accordingly, when formulating the likelihood function for such a road section, one should not assume that the κ's remain constant over time. However, neither should one assume that the change in the κ's with 'y' is entirely haphazard and unpredictable. I will assume that the change in the κ's from one year to the next has something to do with changes in traffic flow on that road section, with changes in precipitation in the region, with changes in driver demography in the state, etc. More specifically, I will **assume that the same regularity that is captured by the model equation** can be exploited to impose some discipline on the time series of κ's for entity 'i' in Equation 12.4. Specifically, the change in $\kappa_{i,y}$ is assumed to be in accord with the following rule:

$$\frac{\kappa_{i,y}}{\kappa_{i,1}} = \frac{\textit{model equation of entity i for year y}}{\textit{model equation of entity i for year 1}} \equiv C_{i,y} \qquad \ldots 12.5$$

$$\textit{that is,} \qquad \kappa_{i,y} = \kappa_{i,1} C_{i,y}$$

[1] As has been discussed in Section 12.2, the use of parameters α_y reflects the belief that the factors they represent (weather, economic conditions etc.) change from year to year. However, the implicit assumption is that this change is similar for all entities in the reference population and that the safety effect of a specified change in these factors is also similar for all entities.

This rule defines[1] the $C_{i,y}$. Note that since the model equation is a function of covariate values and parameters, the definition of the $C_{i,y}$ makes it also a function of covariates and model parameters. At this point a numerical example may be useful.

Numerical Example 12.1. Calculation of $C_{i,y}$.

Consider road section 'i' that is 3.7 km long and on which the AADTs in four consecutive years were 7600,7400,7200 and 6800 vpd. Suppose that model Equation 12.2 is to be used with α_1=0.007742, α_2=0.006659, α_3=0.007066, α_4=0.008256 and β=0.6538. Thus, the mean of the κ's for the reference population for this road section in year 1 is $E\{\kappa_{i,1}\}=3.7\times 0.007742\times 7600^{0.6538}=9.87$ accidents/year. For years 2, 3 and 4, the corresponding values are 8.34, 8.70, and 9.79 accidents/year. Therefore $C_{i,1}=1.00$, $C_{i,2}=8.34/9.87=0.85$, $C_{i,3}=0.88$ and $C_{i,4}=0.99$. Thus, by the rule in Equation 12.5, we are assuming that, whatever the $\kappa_{i,1}$ (the expected accident count of entity 'i' in year 1), in years 2, 3 and 4 it is 0.85, 0.88 and 0.99 of $\kappa_{i,1}$. This change over the four years is a reflection of the change in the AADTs and in parameters α.

The rule captured by Equation 12.5 is a very specific one. The underlying assumption is that how the κ of an entity changes in time can be adequately represented by the change in the covariates over time, and by the model equation. In addition, one is assuming that the model equation which, at best, applies to a set of entities on the average, also holds for each specific entity. One cannot hope that these assumptions are entirely correct. In their defense one can muster two arguments. First, that the assumptions on which Equation 12.5 rests are almost certainly better than what has been done tacitly throughout Part II of this monograph. Namely, to assume that the κ of an entity does not change from year to year in spite of the fact that the covariate or and parameter values do. The second argument is a bit frivolous; I cannot think at this time of a more reasonable rule.

The net effect of the rule in Equation 12.5 is to make all $\kappa_{i,y}$ a function of $\kappa_{i,1}$. By that the number of unknown parameters in the likelihood function is dramatically reduced. The next step is to rid the likelihood function of the remaining unknown parameters $\kappa_{1,1}, \kappa_{2,1}, \ldots, \kappa_{R,1}$. For this, another set of assumptions is needed. The added assumptions are the same as in Section 11.4.

---

[1] I have chosen to place the κ for year 1 in the denominator. This makes $C_{i,1}=1$. The advantage of this choice is to preserve the natural sequence of counting years, the most distant past being year 1. The disadvantage is that, later, when it comes to estimation, the κ for the most distant past is estimated first. I could have chosen to place $\kappa_{i,Y}$ in the denominator. This would have made the C for the last year before treatment equal to one and have the attraction of estimating the most recent κ first. However, it would have forced me to count years in two directions from the time of treatment. Naturally, the end results are not affected by this choice.

Namely that the $\kappa_{1,1}, \kappa_{2,1}, \ldots, \kappa_{R,1}$ are Gamma distributed (see Equation 11.5) with parameters a_i and 'b' as in Equations 11.4. Specifically, that

$$f(\kappa_{i,1}) = \frac{a_i^{\,b}\,\kappa_{i,1}^{\,b-1}\,e^{-a_i\kappa_{i,1}}}{\Gamma(b)} \qquad \ldots 12.6$$

$$\textit{where} \quad a_i = \frac{E\{\kappa_{i,1}\}}{VAR\{\kappa_{i,1}\}} \quad , \quad b = \frac{(E\{\kappa_{i,1}\})^2}{VAR\{\kappa_{i,1}\}} \quad .$$

To eliminate the $\kappa_{1,1}, \kappa_{2,1}, \ldots, \kappa_{R,1}$ from the likelihood function only the parameter 'b' needs to be added to it. How this is done is shown in the 'Derivations'.

Using these three assumptions (about Poisson distributed accident counts, about how the κ's of a certain entity change over the years, and about the distribution of the $\kappa_{i,1}$ in the imaginary reference population) I show in the derivations that the natural logarithm of the likelihood function (divided by a constant) can be written as:

$$\ln(\mathcal{L}/constant) = \sum_{i=1}^{R} \Big(\Big[\sum_{y=1}^{Y+Z} K_{i,y}\ln(C_{i,y})\Big] + b\ln(b/E\{\kappa_{i,1}\}) - \Big(\sum_{y=1}^{Y+Z} K_{i,y} + b\Big)\ln\Big(b/E\{\kappa_{i,1}\} + \sum_{y=1}^{Y+Z} C_{i,y}\Big) + \ln(b) + \ln(b+1) + \ldots + \ln\Big(b + \sum_{y=1}^{Y+Z} K_{i,y} - 1\Big) \Big) \qquad \ldots 12.7$$

The loop is now closed. Equation 12.7 gives the natural logarithm (divided by a constant) of the probability of observing the R×(Y+Z) accident counts $\{K_{i,y}\}$. Once the model equation is chosen, the $C_{i,y}$ and $E\{\kappa_{i,1}\}$ are functions of the known covariate values and of the model parameters. Therefore $\ln(\mathcal{L}/\text{constant})$ is also a function of only the known covariate values and of the unknown model parameters (including the parameter 'b'). Inasmuch as the likelihood function is used here only to find the model parameter values that make it largest, all factors that do not depend on the unknown parameters have been lumped into a constant of no consequence. That set of model parameters which makes $\ln(\mathcal{L}/\text{constant})$ largest will be the set of 'maximum likelihood estimates'. These are then to be used in the model equation to estimate $E\{\kappa_{i,y}\}$. In addition, once $E\{\kappa_{i,y}\}$ has been estimated, one can obtain an estimate of $VAR\{\kappa_{i,y}\}$ since, from Equation 12.6, $VAR\{\kappa_{i,y}\}=(E\{\kappa_{i,y}\})^2/b$.

The identification of the set of parameters that maximizes $\ln(\mathcal{L}/\text{constant})$ is not a trivial task. However, methods and software for finding maxima of complicated functions are available and the matter will not be pursued here. Also, if one does not insist on the EB formulation of the model equation and the rule of Equation 12.5, parameters can be estimated using a statistical software package such as GLIM, SAS, SPSS and the like. The next section is an extended numerical example of how the value of the likelihood function is computed from data and how it is used to find the most likely set of unknown parameters. Following that I will return to the main task, the task of estimating the $\kappa_{i,y}$.

Derivations.

Equation 12.4.

The probability of observing a certain accident count on a specified entity and given year is specified in Equation 12.3. This probability is conditional on the value of the parameter of the Poisson distribution ($\kappa_{i,y}$). The probability of observing another accident count on an entity (say, entity j) and in a certain year (say, year z) is conditional on another parameter of the Poisson distribution ($\kappa_{j,z}$). As long as they are conditional on the values of their parameters, the two accident counts can be taken as statistically independent, even if there is some relationship between $\kappa_{i,y}$ and $\kappa_{j,z}$. Therefore, by the multiplication rule for the probabilities of independent events, the probability of obtaining the set of accidents counts $\{K_{i,y}\}$ when 'i' takes on the values 1, 2, . . . , R and 'y' takes on the values 1, 2, . . . , Y+Z is the double product in Equation 12.4.

Equation 12.7.

It is simpler to deal initially only with entity 'i'. Using the definitional relationship $\kappa_{i,y}=\kappa_{i,1}C_{i,y}$ of Equation 12.5 in Equation 12.4 we obtain:

$$
\begin{aligned}
&P(Accident\,Counts\ \{K_{i,y}\} \mid \kappa_{i,1}, model\ parameters, covariates) \\
&= \prod_{y=1}^{Y+Z} \kappa_{i,y}^{K_{i,y}} e^{-\kappa_{i,y}} / K_{i,y}! = \prod_{y=1}^{Y+Z} (C_{i,y}\kappa_{i,1})^{K_{i,y}} e^{-C_{i,y}\kappa_{i,1}} / K_{i,y}! \\
&= \left[\prod_{y=1}^{Y+Z} \frac{C_{i,y}^{K_{i,y}}}{K_{i,y}!}\right] \left(\kappa_{i,1}^{\sum_{y=1}^{Y+Z} K_{i,y}} e^{-\kappa_{i,1}\sum_{y=1}^{Y+Z} C_{i,y}}\right) \qquad \ldots 12.8
\end{aligned}
$$

By this, all the $\kappa_{i,y}$, except for $\kappa_{i,1}$, have been replaced by the covariates and parameters of the model equation. These are embedded in the $C_{i,y}$. It remains to get rid of $\kappa_{i,1}$. Here is where the Empirical Bayes element comes in. The probability in Equation 12.8 is conditional on the $\kappa_{i,1}$. To remove the condition, one has to state what is the probability of $\kappa_{i,1}$, then to include this probability in the right-hand side of Equation 12.7, and finally to integrate over all possible values $\kappa_{i,1}$. As in Chapter 11, the two-parameter gamma distribution is chosen to represent the probability distribution of $\kappa_{i,1}$. This is done for two reasons. First, the choice brings forth a convenient result after integration. Second, the gamma distribution can well approximate the probability distribution of a random variable that takes on only positive values, provided that it has a single-peak and long tail to the right. The

Gamma probability density function of $\kappa_{i,1}$ and the relationship of its parameters to mean and variance was given by Equations 12.6:

$$f(\kappa_{i,1}) = \frac{a_i^{\,b}\,\kappa_{i,1}^{\,b-1}\,e^{-a_i\kappa_{i,1}}}{\Gamma(b)}$$

$$where \quad a_i = \frac{E\{\kappa_{i,1}\}}{VAR\{\kappa_{i,1}\}} \quad , \quad b = \frac{(E\{\kappa_{i,1}\})^2}{VAR\{\kappa_{i,1}\}} \quad . \qquad \ldots 12.6$$

The parameter 'b' has no subscript because in previous work it has been found that estimates $VAR\{\kappa_{i,y}\}$ and $E\{\kappa_{i,y}\}$ are often systematically associated and that for all 'y' and 'i' $VAR\{\kappa_{i,y}\}=[E\{\kappa_{i,y}\}]^2/b$. Using the probability in Equation 12.8 in conjunction with Equation 12.6, and integrating over all $\kappa_{i,1}$ we get:

$$P(Count = K_{i,1},K_{i,2},\ldots,K_{i,Y+Z}|\ model\ parameters,\ covariates)$$

$$=[\prod_{y=1}^{Y+Z}\frac{C_{i,y}^{K_{i,y}}}{K_{i,y}!}][\frac{a_i^{\,b}}{\Gamma(b)}]\int_{\kappa_{i,1}=0}^{\infty}\kappa_{i,1}^{(\sum_{y=1}^{Y+Z}K_{i,y})+b-1}\,e^{-\kappa_{i,1}(a_i+\sum_{y=1}^{Y+Z}C_{i,y})}\,d\kappa_{i,1}$$

$$=[\prod_{y=1}^{Y+Z}\frac{C_{i,y}^{K_{i,y}}}{K_{i,y}!}][\frac{a_i^{\,b}}{\Gamma(b)}][\frac{\Gamma(\sum_{y=1}^{Y+Z}K_{i,y}+b)}{(a_i+\sum_{y=1}^{Y+Z}C_{i,y})^{\sum_{y=1}^{Y+Z}K_{i,y}+b}}] \qquad \ldots 12.9$$

$$=[\prod_{y=1}^{Y+Z}\frac{C_{i,y}^{K_{i,y}}}{K_{i,y}!}][\frac{a_i^{\,b}}{(a_i+\sum_{y=1}^{Y+Z}C_{i,y})^{\sum_{y=1}^{Y+Z}K_{i,y}+b}}][b(b+1)(b+2)\ldots(b+\sum_{y=1}^{Y+Z}K_{i,y}-1)]$$

When $\Sigma K_{i,y}=0$, the product in the last square brackets is 1, when $\Sigma K_{i,y}=1$, the product is b and so on. In this use has been made of the definite integral

$$\int_0^{\infty} x^n e^{-\gamma x}dx = \frac{\Gamma(n+1)}{\gamma^{n+1}}$$

The product of the expressions in Equation 12.9, one for each entity in the data set, is the sought likelihood function, $\mathcal{L}$. Inasmuch as it is used here only to find parameter values that make $\mathcal{L}$ largest, all factors that do not depend on the unknown parameters can be lumped into one constant the composition of which is immaterial for our purposes. In addition we replace a_i by $b/E\{\kappa_{i,1}\}$ based on Equations 12.6. With these changes Equation 12.9 takes on the form:

$$\mathcal{L} = constant \times \prod_{i=1}^{R} \left(\left[\prod_{y=1}^{Y+Z} (C_{i,y})^{K_{i,y}}\right] \times \left[\frac{(b/E\{\kappa_{i,1}\})^{b}}{(b/E\{\kappa_{i,1}\} + \sum_{y=1}^{Y+Z} C_{i,y})^{\sum_{y=1}^{Y+Z} K_{i,y}+b}} \right] \times \left[b(b+1)(b+2)\ldots(b+\sum_{y=1}^{Y+Z} K_{i,y}-1) \right] \right) \quad \ldots 12.10$$

What parameter values maximize $\mathcal{L}$, also maximize the natural logarithm of ($\mathcal{L}$/constant). Therefore, the maximum likelihood estimates of the parameters are those that maximize:

$$\ln(\mathcal{L}/constant) = \sum_{i=1}^{R} \left(\left[\sum_{y=1}^{Y+Z} K_{i,y} \ln(C_{i,y})\right] + b\ln(b/E\{\kappa_{i,1}\}) - \left(\sum_{y=1}^{Y+Z} K_{i,y}+b\right)\ln(b/E\{\kappa_{i,1}\} + \sum_{y=1}^{Y+Z} C_{i,y}) + \ln(b) + \ln(b+1) + \ldots + \ln(b+\sum_{y=1}^{Y+Z} K_{i,y}-1) \right)$$

This is Equation 12.7.

12.4 AN ILLUSTRATION

A numerical example may help to make all this more tangible. Suppose that we have four years of data for six rural two-lane road sections as listed in Table 12.1 and shown in Figure 12.3. The task is to illustrate how the log-likelihood of Equation 12.7 is computed for a given set of parameter values, and how the parameter values that make it largest can be identified.

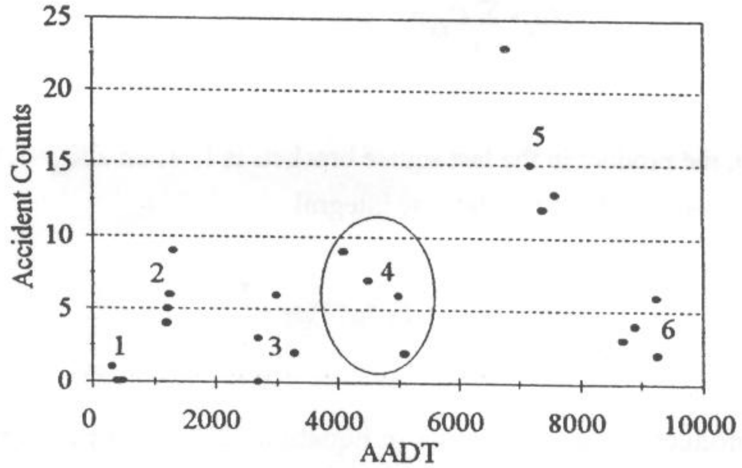


Figure 12.3. Accident counts versus AADT.

Table 12.1. Data for six road sections.

Road Section	Year	AADT	Length in km	Accident counts	Road Section	Year	AADT	Length in km	Accident counts
1	1	320	3.1	1	4	1	4100	5.6	9
1	2	400	3.1	0	4	2	4500	5.6	7
1	3	450	3.1	0	4	3	5000	5.6	6
1	4	500	3.1	0	4	4	5100	5.6	2
2	1	1220	4.2	5	5	1	7600	3.7	13
2	2	1200	4.2	4	5	2	7400	3.7	12
2	3	1300	4.2	9	5	3	7200	3.7	15
2	4	1250	4.2	6	5	4	6800	3.7	23
3	1	3300	2.7	2	6	1	9250	1.9	6
3	2	3000	2.7	6	6	2	9260	1.9	2
3	3	2700	2.7	0	6	3	8700	1.9	3
3	4	2700	2.7	3	6	4	8900	1.9	4

In this data set the accident counts of each road section are seen to form a cluster. Thus, e.g., the four accident counts inside the ellipse are for road section 4, one point for each year. In parameter estimation , one may not regard these four points as if they were representing for four separate road sections. Nor would it be realistic to represent them as a road section on which the conditions remained constant since flow changed over time. The rule in Equation 12.5 seems to be a reasonable compromise and the likelihood function in Equation 12.7 is its representation. One cannot hope to discern in Figure 12.3 a clear relationship between the expected accident counts and traffic. This is partly so because the sections differ not only in traffic but also in length. Thus, e.g., road section 5 is longer than road section 6. Also, the underlying relationship is masked by the random variation of accident counts ($K_{i,y}$) around their means ($\kappa_{i,y}$). In addition, as shown in Figure 12.2, it is not the means behind the accident counts (the $\kappa_{i,y}$) of these six road sections that are on the curve represented by the model equation; it is the expected accident counts of the reference populations of such road sections (the $E\{\kappa_{i,y}\}$) that form the curve.

To compute a starting value for the log-likelihood in Equation 12.7 we need to choose a model equation and to make an initial guess about its parameters. In this example I will use the form of Equation 12.2, namely:

$$E\{\kappa_{i,y}\}=d_i\alpha_y F_{i,y}{}^{\beta}$$

This means that there are five unknown model parameters ($\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta$) and the parameter 'b' which specifies the VAR$\{\kappa_{i,1}\}$ (see Equation 12.6). As an initial guess I will take $\hat{\beta}=1$. Were this true, accident frequency would be proportional to AADT. Also, as a point of departure I will take all four α's to be the same. As a starting guess, since the 5.1 km long section 4 had an average of 6 accidents/year with an AADT of about 5000, I will use $\hat{\alpha}= 6/(5.1\times5000)$ which is approximately 0.0002. Finally, I will start with $\hat{b}=1$, guessing that in this case the VAR$\{\kappa_{i,1}\}=(E\{\kappa_{i,1}\})^2$. Now the estimates of $E\{\kappa_{i,y}\}$ and of $C_{i,y}$ can be computed as is shown in Table 12.2. Thus, e.g., with the current parameter values, $\hat{E}\{\kappa_{1,1}\}=3.1\times0.0002\times320=0.198$ accidents while $\hat{E}\{\kappa_{1,4}\}= 3.1\times0.0002\times500$ $=0.310$ accidents. It follows from the definition of $C_{i,y}$ (Equation 12.5) that $\hat{C}_{1,4}=0.31/0.20=1.56$.

Table 12.2. Starting values of $\hat{E}\{\kappa_{i,y}\}$ and $\hat{C}_{i,y}$.

Section	Year	$\hat{E}\{\kappa_{i,y}\}$	$\hat{C}_{i,y}$	$\Sigma\hat{C}_{i,y}$	Section	Year	$\hat{E}\{\kappa_{i,y}\}$	$\hat{C}_{i,y}$	$\Sigma\hat{C}_{i,y}$
1	1	0.20	1.00		4	1	4.59	1.00	
1	2	0.25	1.25		4	2	5.04	1.10	
1	3	0.28	1.41		4	3	5.60	1.22	
1	4	0.31	1.56	5.22	4	4	5.71	1.24	4.56
2	1	1.02	1.00		5	1	5.62	1.00	
2	2	1.01	0.98		5	2	5.48	0.97	
2	3	1.09	1.07		5	3	5.33	0.95	
2	4	1.05	1.02	4.07	5	4	5.03	0.89	3.82
3	1	1.78	1.00		6	1	3.52	1.00	
3	2	1.62	0.91		6	2	3.52	1.00	
3	3	1.46	0.82		6	3	3.31	0.94	
3	4	1.46	0.82	3.55	6	4	3.38	0.96	3.90

With these values ready, we can now compute each of the four summands in Equation 12.7 for each road section, and then their total. This is done in Table 12.3. Thus, e.g., for road section 2, the value of the first summand is $5\times\ln(1.00)+4\times\ln(0.98)$ $+9\times\ln(1.07)+6\times\ln(1.02)=0-0.08+0.61+0.11=0.647$. The second summand is $b\times\ln(1/1.02)=-0.02$. The third summand is $-(5+4+9+6+1)\times\ln(1/1.02+4.07)=-40.48$. The fourth summand is $\ln(1)+\ln(2)+\ldots\ln(1+24-1)=$ 54.78. The row sum (14.93) is the contribution of road section 2 to the total log-likelihood in the bottom cell (137.48).

Table 12.3. Values for the four components of the log-likelihood function.

	I	II	III	IV	
Section i	$\Sigma_{y=1}^{4}$ $[K_{i,y}\ln(C_{i,y})]$	$b\times\ln(b/E\{\kappa_{i,1})$	$-(\Sigma_{y=1}^{4} K_{i,y}+b)\times$ $\ln(b/E\{\kappa_{i,1}\}+\Sigma_{y=1}^{4} C_{i,y})$	$\ln(b)+\ln(b+1)+\ldots$ $+\ln(b+\Sigma_{y=1}^{4}K_{i,y}-1)$	Row sum for section i
1	0.00	1.62	-4.66	0	-3.04
2	0.65	-0.02	-40.48	54.78	14.93
3	-1.17	-0.58	-16.95	17.50	-1.20
4	2.28	-1.52	-39.10	54.78	16.43
5	-3.69	-1.73	-88.62	201.01	106.97
6	-0.34	-1.26	-22.92	27.90	3.39
				Sum of row sums = ln($\mathcal{L}$/constant) =	137.48

Was I to change $\hat{\beta}$ from 1 to 1.1, the log-likelihood would increase to 139.61. This is a very sizeable increase, remembering that an increase of the natural logarithm by 1 amounts to multiplication by a factor of 2.72. A further increase of $\hat{\beta}$ to 1.2, will cause a drop of the log-likelihood to 138.12. Thus, keeping all other parameters unchanged, there is a local maximum near $\hat{\beta}$=1.1. One can drive the log-likelihood further up by allowing all the other parameters to change. Thus, making $\hat{\beta}$=1.09 and changing the other parameters to $\hat{\alpha}_1$=0.00022, $\hat{\alpha}_2$=0.00019, $\hat{\alpha}_3$=0.00020, $\hat{\alpha}_4$=0.00024 and $\hat{b}$=2.64 the log-likelihood increases to 141.09. No small change of the parameters would help to increase log-likelihood further. Thus, we have a **local maximum**. Had I started my hill climbing from a different initial guess, I might have reached a different peak of the log-likelihood function. Indeed, starting from the initial guesses $\hat{\beta}$=0.7, $\hat{\alpha}$=0.003 and $\hat{b}$=2 I found another local maximum at $\hat{\alpha}_1$=0.0032, $\hat{\alpha}_2$=0.0027, $\hat{\alpha}_3$=0.0028, $\hat{\alpha}_4$=0.0034 and $\hat{\beta}$=0.763 and $\hat{b}$=3.8. Here the value of the log-likelihood function is 142.33.

It is now apparent that it is relatively simple to find a local maximum of the log-likelihood function by changing the parameter values one after another till no further improvement is possible. However, the likelihood function has multiple local maxima. There is no assurance that a certain starting guess will lead to the highest peak of the likelihood function. Therefore, to find its globally largest value, and thereby to identify the set of most likely parameters, is not simple. One has to use some systematic search algorithm that does not get trapped in the vicinity of a local maximum. There exist tried and tested algorithms for this purpose. I have been using the Nelder-Mead algorithm (see, e.g., Nash, 1990). Applying this algorithm to our data results in estimates $\hat{\alpha}_1$=0.00774, $\hat{\alpha}_2$=0.00666, $\hat{\alpha}_3$=0.00707, $\hat{\alpha}_4$=0.00826, $\hat{\beta}$=0.654 and $\hat{b}$=3.86 where the value of the log-likelihood function is 142.43. While the improvement in the value of the likelihood function is only about 11% (142.43-142.33=0.1 and $e^{0.1}$=1.11) the change in the parameter values appears to be large. An attractive option for parameter estimation is to use a statistical software such as GLIM, SAS, SPSS or any other software for generalized linear modeling. In the present case, the use of GLIM (with the Negative Binomial error structure) yields only a slightly smaller log-likelihood value (143.40) than the Nelder-Mead search. However, the GLIM algorithm is fast, provides useful statistics (standard errors of estimates) and once mastered the software is easy to use. Note, that the likelihood function in Equation 12.7 is based on the rule in Equation 12.5 and therefore accounts reasonably for how each entity in the data set changes over time. This does not seem true for GLIM and similar software[1]. In consequence, use of the likelihood function in Equation 12.7 will usually produce somewhat different parameter estimates than GLIM and similar statistical software. The four sets of estimates mentioned so far are shown in Table 12.4.

Table 12.4. Four sets of parameter estimates.

	First local maximum	Second local maximum	Nelder-Mead	GLIM
$\hat{\alpha}_1$	0.00022	0.0032	0.00774	0.00547
$\hat{\alpha}_2$	0.00019	0.0027	0.00666	0.00469
$\hat{\alpha}_3$	0.00020	0.0028	0.00707	0.00490
$\hat{\alpha}_4$	0.00024	0.0034	0.00826	0.00563
$\hat{\beta}$	1.090	0.763	0.654	0.701
$\hat{b}$	2.64	3.80	3.86	3.86
$\ln \mathcal{L}$	141.09	142.33	142.43	142.40

The corresponding $\hat{E}\{\kappa\}$ for a range of AADTs is shown in Figure 12.4.

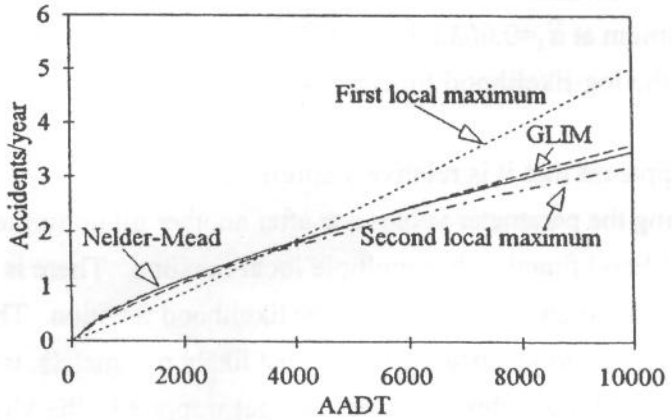


Figure 12.4. Comparison of fitted values by four sets of parameters.

[1] Maher and Summersgill (1996) write that "... studies typically obtained accident data over a four or five year period (the same period for all sites) and generally models were fitted using as the unit of data the total accidents over the whole period." They note that if each year for an entity is to be a unit of data then the "...terms can not be set out in additive linear form." This is why they suggest an iterative approach .

Several observations follow. First, parameters from a local maximum can produce fitted values that are substantively different from those produced by the parameters near the global maximum. Therefore, it is essential to search for the global maximum. Second, that even when the estimated parameters seem substantively different (as in the three right columns of Table 12.4), the corresponding fitted values may be very similar. Naturally, the larger the exponent β, the lesser must the α's be. Third, that the magnitude of the log-likelihood and parameter 'b' go hand in hand. The better the fit to the data, the larger the log-likelihood. Simultaneously, the better the fit to the data, the lesser will be $VAR\{\kappa\}$ and therefore the larger will be $b=E\{\kappa\}^2/VAR\{\kappa\}$.

It may appear now that in the last three sections I have strayed far from the main task of interpreting observational Before-After studies. The project of Chapter 12 was to put on a firmer footing the prediction of what would have been the safety of a treated entity in the 'after' period, had it been left without treatment. To remove the impediments to better prediction we aimed to do away with the fixed-duration 'before' period and to use accident counts from the more distant past. We also wanted to estimate the time series $\kappa_1, \kappa_2, \ldots, \kappa_Y$ for the 'before' period. All this focusses on the treated entities. Yet Sections 12.2 to 12.4 dealt with fitting a model to a reference population of untreated entities. As has been shown in Figure 12.1, this diversion is of essence. To make progress toward better prediction I need a multivariate model for the reference populations. This is why I had to clarify some matters about multivariate statistical models. I had to declare what is being estimated by a multivariate model (Equation 12.2); I had to explain what the model equation is and what the many assumptions it represents; and I had to develop the likelihood function and show how the parameters of the model equation are to be estimated. We are now ready to return to the main task. The next question is how to estimate the time series $\kappa_1, \kappa_2, \ldots, \kappa_Y$ for a treated entity and how to predict for it the time series $\kappa_{Y+1}, \kappa_{Y+2}, \ldots, \kappa_{Y+Z}$.

12.5 HOW TO ESTIMATE THE $\kappa_1, \kappa_2, \kappa_3, \ldots, \kappa_Y$ FOR SOME ENTITY?

At this point all the preliminary activities associated with the multivariate statistical modeling have been completed. Among other things, we have chosen a model equation and, using data about accident counts and covariates for a set of untreated entities, we have estimated the parameters of the model equation. Therefore for any value of the covariates in the model equation and for any year we can compute estimates of the mean of the κ's, the variance of the κ's, and the ratio C defined by Equation 12.5.

Let 'i' be a certain treated entity. In years y=1, 2, . . . ,Y it had specific covariate values. Using these covariate values we can now compute $\hat{E}\{\kappa_{i,y}\}$, $V\hat{A}R\{\kappa_{i,y}\}$ and $\hat{C}_{i,y}$ for the imagined reference population of this entity. Now we can tackle the next task, namely:

Estimate the $\kappa_{i,1}$, $\kappa_{i,2}$, $\kappa_{i,3}$, . . . ,$\kappa_{i,Y}$ of entity 'i' that has been treated at the end of year Y . This entity has recorded $K_{i,y}$ accidents in years y=1, 2, . . . ,Y, and, by its traits (covariates), belongs to an imagined reference populations with $E\{\kappa_{i,y}\}$, $VAR\{\kappa_{i,y}\}$ and $C_{i,y}$.

For practical and didactic reasons I will discuss this estimation task in two stages. In the first stage I will free estimation from the constraint which now compels us to use short 'before' periods and to pretend that in these two or three 'before' years κ was constant. The result of the first stage will be a method for estimating $\kappa_{i,y}$ for any year 'y' using the entire accident history $K_{i,1}$, $K_{i,2}$, . . . , $K_{i,Y}$. In the second stage I will recast estimation into the EB framework. This will once again remove from estimation the threat of RTM bias and will allow statements to be made about the probability distribution of the estimated values $\hat{\kappa}_{i,y}$.

a. Stage 1: Maximum Likelihood Estimation

Much of the groundwork has been laid in Section 12.3. I have argued there that while the κ of an entity changes from one year to another, there must be some discipline and orderliness to this change. This discipline and order reflect facts such as that a road section does not change its alignment every year nor does driver demography change abruptly. Important features of entities either remain largely unchanged over time, change slowly, or change predictably. As in Equation 12.5, I will again assume that the change in the κ's of an entity follows the rule: $\kappa_{i,y}=\kappa_{i,1}C_{i,y}$. In this, $C_{i,y}$ is estimated by the ratio of the fitted model equation for year 'y' divided by that for year 1. This rule led to Equation 12.8 that states the conditional probability of obtaining the accident counts $K_{i,1}$ $K_{i,2}$, . . . , $K_{i,Y}$ given that $\kappa_{i,1}$ is the expected value for this entity in year 1. I show in the derivations at the end of this section, that when Equation 12.8 is construed to be the likelihood function for $\kappa_{i,1}$, its maximum likelihood estimate is given by Equation 12.11.

$$\hat{\kappa}_{i,1} = \frac{\sum_{y=1}^{Y} K_{i,y}}{\sum_{y=1}^{Y} \hat{C}_{i,y}} \qquad \ldots 12.11$$

Once we have an estimate of $\kappa_{i,1}$, the rule in Equation 12.5 allows estimation of the remaining $\kappa_{i,y}$ by:

$$\hat{\kappa}_{i,y} = \hat{\kappa}_{i,1}\hat{C}_{i,y} \qquad \ldots 12.12$$

There is nothing unique about $\kappa_{i,1}$ for which it has to be estimated first. This is but the reflection of the convention I have chosen in Equation 12.5. Had I chosen there to place $\kappa_{i,Y}$ in the denominator of Equation 12.5 instead of $\kappa_{i,1}$, the values of the C's would be different and Equation 12.11 would be for $\hat{\kappa}_{i,Y}$. However, all numerical results would remain the same.

Equation 12.11 is a comforting result. First, it is consistent with what is well known and usually done. To see this, consider the following line of reasoning. Were nothing to change from one year to the next, the model equation would have the same value for all years, and therefore the $C_{i,y}$ would be 1 for all 'y'. In this case the estimate of $\kappa_{i,1}$ in Equation 12.11 would be the usual - sum of accident counts divided by the number of years. Thus, what we found is consistent with what is done (usually) when one is assuming (incorrectly) that the covariate values do not change in time and therefore κ_i is constant over time. Second, one of the main objectives of Part III was to free estimation from the unrealistic assumption that nothing changes from one year to the next. Equations 12.11 and 12.12 attain this objective. The estimate of each $\kappa_{i,y}$ is now a function of the entire set $\{K_{i,1}, K_{i,2}, \ldots, K_{i,Y}\}$ of accident counts and also of the changing conditions as reflected by the set $\{C_{i,1}=1, C_{i,2}, \ldots, C_{i,Y}\}$. Conveniently, the estimator in Equation 12.11 happens to be a particularly simple function. The following numerical example serves for illustration.

Numerical Example 12.2. Estimating the $\kappa_{i,y}$ of a road section.

During the 8 years 1975 to 1982 a 1.6 miles long undivided two-lane rural road in New York State recorded the set of {1, 4, 5, 1, 4, 1, 3, 0} non-intersection accidents. This road section will be entity 'i'. The task is to estimate the $\kappa_{i,y}$ for this road section for the years 1975-1982. Parameters of a multivariate model, of the kind that has been specified by Equation 12.2, for non-intersection accidents on two-lane undivided rural roads in New York State have been estimated. The $\hat{\beta}$=0.776 and the $\hat{\alpha}$'s are listed in Table 12.4. In this table, y=1 corresponds to 1975 and Y=8 to 1982. Data about the corresponding traffic flows are given in column 3 of Table 12.5. Using these one can compute $\hat{E}\{\kappa_{i,y}\}$ by Equation 12.2 and $\hat{C}_{i,y}$ by Equation 12.5.

Table 12.5. Computation of $\hat{E}\{\kappa_{i,y}\}$ and $\hat{C}_{i,y}$.

y	$\hat{\alpha}_y$	$AADT_y$	$\hat{E}\{\kappa_{i,y}\}$	$\hat{C}_{i,y}$
1	0.002844	1199	1.115	1.000
2	0.002885	1201	1.132	1.016
3	0.002745	1175	1.059	0.950
4	0.002550	1163	0.976	0.876
5	0.002662	1098	0.975	0.874
6	0.002634	1042	0.926	0.831
7	0.002479	1038	0.869	0.779
8	0.002699	1038	0.946	0.849
Sum				7.173

Numerical Example 12.2. (continued).

For example, in 1975 (y=1), for the imagined set of road sections of this kind, with AADT=1199, by Equation 12.2, $\hat{E}\{\kappa_{i,1}\}=1.6\times0.002844\times1199^{0.776}=1.115$ non-intersection accidents. In 1976, general conditions changed from those prevailing in 1975, resulting in a slightly larger α. In addition, AADT increased slightly. Therefore $\hat{E}\{\kappa_{i,2}\}=1.132$ non-intersection accidents making $\hat{C}_{i,2}=1.132/1.115=1.016$. Now the question about the κ's of this road section can be answered. The sum of accident counts is 19 and the sum of $\hat{C}$'s is 7.173.

Now we can estimate κ's. By Equation 12.11, for 1975, $\hat{\kappa}_{i,1}=19/7.173=2.65$ non-intersection accidents. By Equation 12.12, in 1976, $\hat{\kappa}_{i,2}=2.65\times1.016=2.69$ non-intersection accidents while, say, in 1981, $\hat{\kappa}_{i,7}=2.65\times0.779=2.06$ non-intersection accidents.

A few aspects of this numerical example require comment. Were one to simply average the eight accident counts, $\kappa_{i,1}$ would be estimated as 19/8=2.37 accidents/year (instead of 2.65 in the numerical example). The difference between the two estimates is due to the denominator in Equation 12.11. In the example I have used $\Sigma C_y=7.173$ in the denominator instead of 8 in the simple average[1]. This difference may be thought too small to deserve so much attention. Indeed, the point is not that a more appropriate but only slightly different estimate of $\kappa_{i,1}$ has been obtained. The main points are two. First, when calculating the simple average for a number of years, one is speaking about an average over the entire period, without any ability to differentiate between the κ's that may have prevailed in each year within that period. In contrast, using Equations 12.11 and 12.12 each year has its own κ. Second, one usually hesitates to estimate the $\kappa_{i,y}$ using more than, say, three years of accident counts. The fear is that by using accident counts that are too remote in time from the period of interest the integrity of the estimate may be compromised. The principal merit of the estimators in Equations 12.11 and 12.12 is to make it possible to use accident counts from many years to estimate the $\kappa_{i,y}$ in a specific year, provided that the change in κ's over the years can be accounted for by the coefficients $C_{i,y}$. The $C_{i,y}$ are available for estimation because of the multivariate model for the reference populations. The increase in estimation accuracy is not so much due to the modifying influence of the $C_{i,y}$'s, but is attributable mainly to the increase in the number

[1] To explain at the intuitive level why in this case ΣC is less that 8, note that in Table 12.5, for all y>2, the α's and the AADT's are less than those for y=1. Had α_1 and $AADT_1$ prevailed in the years y>2, one would expect to record somewhat larger accident counts than what actually went into the sum in the numerator. The diminished denominator compensates for this shortfall.

of accident counts that, by virtue of the $C_{i,y}$, can now be legitimately brought to bear upon the estimation task.

To illustrate, a common way to estimate $\kappa_{i,y}$ for a certain year would be compute the average for the neighboring years. Recall that in Numerical Example 12.2 the accident counts were {1, 4, 5, 1, 4, 1, 3, 0}. If so, by a simple average we would estimate $\kappa_{i,2}$ to be (1+4+5)/3=3.33 accidents/year and $\kappa_{i,7}$ as (1+3+0)/3=1.33 accidents/year. However, the expected accident frequency of a road does not usually change so much. I suspect that the large difference between 3.33 and 1.33 reflects more random variation than change in the underlying expected frequency. Comparing these averages with the maximum likelihood estimates obtained earlier (2.69 and 2.06) the difference is obvious. Naturally, estimates based on 8 years of accident counts bound to be less given to random fluctuation than estimates based on only 3 years of accident counts.

b. Empirical Bayes Estimation

In Section 11.1 I spoke of the 'shaky foundation' - the assumption that by using accident counts alone one can obtain sensible estimates of $\kappa_{i,y}$. Specifically, a systematic estimation bias was shown to exist when the accident history of an entity had something to do with the decision to administer treatment. In that section, the existence of this regression-to-mean (RTM) bias has been amply demonstrated both in theory and by empirical evidence. The suggested remedy was the adoption of the Empirical Bayes (EB) approach. Within the EB framework, estimation of $\kappa_{i,y}$ is based on the joint use of two kinds of clues: those contained in accident counts and those contained in the traits of the entity and the corresponding reference population.

Estimation by Equations 12.11 and 12.12 is also subject to the RTM bias. It is true that when accident counts for more years are used for estimation, the magnitude of the bias may be smaller. The means for using more years of accident counts for estimation have just been provided. Still, the danger of the RTM bias exists no matter how long is the accident record used, and it is best to remove it by the EB method. The use of the EB method will be shown to be entirely painless, when a multivariate model for the reference population is available. All that remains to be done is to sort out the algebra.

We wish to estimate $\kappa_{i,y}$, the expected number of accidents of the entity 'i' in year 'y'. To repeat the logic already laid out a few times, we imagine a reference population of entities with the same covariates and the same accident history as entity 'i'. The mean of the κ's in this imagined reference population is the best estimate of the sought $\kappa_{i,y}$ because there is nothing to distinguish entity 'i' from the entities of this imagined reference population. Since $\kappa_{i,y}$ is one of the κ's from this imagined reference population, the variance of the κ's in the imagined reference population can serve as the variance with which $\kappa_{i,y}$ is being estimated. All that needs to be done is to show how to obtain the probability distribution of κ's in this imagined reference population and to determine its mean and its variance.

Already in Section 12.3, I have made use of an imagined reference population for entity 'i' in year 1. At that time it was composed of entities that have exactly the same covariates as entity 'i', but not necessarily the same accident history. The probability density function of this reference population was denoted by $f(\kappa_{i,1})$; it was assumed to be Gamma with parameters $a_i=E\{\kappa_{i,1}\}/VAR\{\kappa_{i,1}\}$ and $b=(E\{\kappa_{i,1}\})^2/VAR\{\kappa_{i,1}\}$ (see Equation 12.6). The same 'Gamma assumption'[1] will be used here.

Consider now a subset of entities from the imagined reference population in which all entities have not only the same covariates as entity 'i' but also the same history of accident counts $K_{i,1}, K_{i,2}, \ldots, K_{i,Y}$. Now the question is:

> **If** the κ's in the original reference population of entity 'i' are Gamma distributed with a probability density function $f(\kappa_{i,1})$ and 'a_i' and 'b' as parameters, **then**, what is the probability density function $f(\kappa_{i,1}|K_{i,1}, K_{i,2}, \ldots, K_{i,Y})$ of the κ's in that subset of the original reference population in which all imagined entities have recorded $K_{i,1}, K_{i,2}, \ldots, K_{i,Y}$ accidents?

I show in the 'derivations' at the end of this section that this new probability density function $f(\kappa_{i,1}|K_{i,1}, K_{i,2}, \ldots, K_{i,Y})$ is also Gamma and given by an equation identical to Equation 12.6, except that the parameter 'a_i' is now replaced by $a_i+\Sigma C_{i,y}$ and the parameter 'b' by $b+\Sigma K_{i,y}$. That is,

$$f(\kappa_{i,1}|K_{i,1},K_{i,2},\ldots,K_{i,Y}) = \frac{(a_i+\sum_{y=1}^{Y}C_{i,y})^{b+\sum_{y=1}^{Y}K_{i,y}}\ \kappa_{i,1}^{\,b+\sum_{y=1}^{Y}K_{i,y}-1}\ e^{-\kappa_{i,1}(a_i+\sum_{y=1}^{Y}C_{i,y})}}{\Gamma(b+\sum_{y=1}^{Y}K_{i,y})} \qquad \ldots 12.13$$

Since we know the probability density function, we can find its mean and its variance. The mean serves to estimate the $\kappa_{i,1}$ for road section 'i' with the accident count history $K_{i,1}, K_{i,2}, \ldots, K_{i,Y}$, and the variance estimates the variance of $\hat{\kappa}_{i,1}$. The expressions for the mean $E\{\kappa_{i,1}|K_{i,1}, K_{i,2}, \ldots, K_{i,Y}\}$ and the variance $VAR\{\kappa_{i,1}|K_{i,1}, K_{i,2}, \ldots, K_{i,Y}\}$ are obtained in the 'Derivations' at the end of this section. On this basis, for road section 'i' that has recorded $K_{i,1}, K_{i,2}, \ldots, K_{i,Y}$ accidents we estimate for year 1:

---

[1] The 'Gamma assumption' has been introduced in Section 11.4 and discussed in a footnote there.

$$\hat{\kappa}_{i,1}=\frac{\hat{b}+\sum_{y=1}^{Y}K_{i,y}}{\frac{\hat{b}}{\hat{E}\{\kappa_{i,1}\}}+\sum_{y=1}^{Y}\hat{C}_{i,y}}\ ;\quad V\hat{A}R\{\hat{\kappa}_{i,1}\}=\frac{\hat{b}+\sum_{y=1}^{Y}K_{i,y}}{(\frac{\hat{b}}{\hat{E}\{\kappa_{i,1}\}}+\sum_{y=1}^{Y}\hat{C}_{i,y})^2}=\frac{\hat{\kappa}_{i,1}}{\frac{\hat{b}}{\hat{E}\{\kappa_{i,1}\}}+\sum_{y=1}^{Y}\hat{C}_{i,y}} \qquad \ldots 12.14$$

It remains to show how the $\kappa_{i,y}$ for years 2, 3, . . . ,Y are to be estimated. This is simple. We have assumed earlier that $\kappa_{i,y}=\kappa_{i,1}C_{i,y}$. From this assumption it follows that:

$$\hat{\kappa}_{i,y}=\hat{C}_{i,y}\hat{\kappa}_{i,1} \text{ and } V\hat{A}R\{\hat{\kappa}_{i,y}\}=\hat{C}_{i,y}{}^2V\hat{A}R\{\hat{\kappa}_{i,1}\} \qquad \ldots 12.15$$

With this in the bag I now return to the context of Numerical Example 12.2.

Numerical Example 12.3. EB estimation of $\kappa_{i,y}$.

I use here the data from Numerical Example 12.2. During the 8 years 1975-1982 a 1.6 mile long undivided two-lane road in New York state recorded 19 non-intersection accidents. In addition to the data already given in Numerical Example 12.2, we use here the estimate $\hat{b}$=5.571 obtained during the multivariate model fitting. From Table 12.5, $\hat{E}\{\kappa_{i,1}\}$=1.115 accidents/year. The sum of C's in Table 12.5 is 7.173. With this, by Equations 12.14, $\hat{\kappa}_{i,1}$=(5.571+19)/(5.571/1.115 +7.173)=24.57/12.17=2.02 accidents/year and $[V\hat{A}R\{\hat{\kappa}_{i,1}\}]^{0.5}$= $[2.02/12.17]^{0.5}$=0.41 accidents/year.

From here, using Equation 12.15, estimates of all other κ's and of their standard deviations can be obtained. Thus, e.g., $\hat{\kappa}_{i,2}$=1.016×2.02=2.05 accidents/year with $\hat{\sigma}\{\hat{\kappa}_{i,2}\}$= 1.016×0.41=0.42 accidents/year; $\hat{\kappa}_{i,7}$=0.779×2.02=1.57 accidents/year with $\hat{\sigma}\{\hat{\kappa}_{i,7}\}$= 0.779×0.41= 0.32 accidents/year, etc.

Recall that in Numerical Example 12.2, when using only accident counts, the estimate of $\kappa_{i,1}$ was 2.65 accidents/year. The road sections that make up the imagined reference population of entity 'i' when matched on all covariates but not accident counts have a mean of 1.115 accidents/year. The EB estimate is 2.02 accidents/year. The EB method uses the evidence for the imagined reference population (1.115 accidents/year) jointly with the evidence from the accident count (2.65 accidents/year) to produce an estimate based on both clues (2.02 accidents/year) as shown in Figure 12.5.

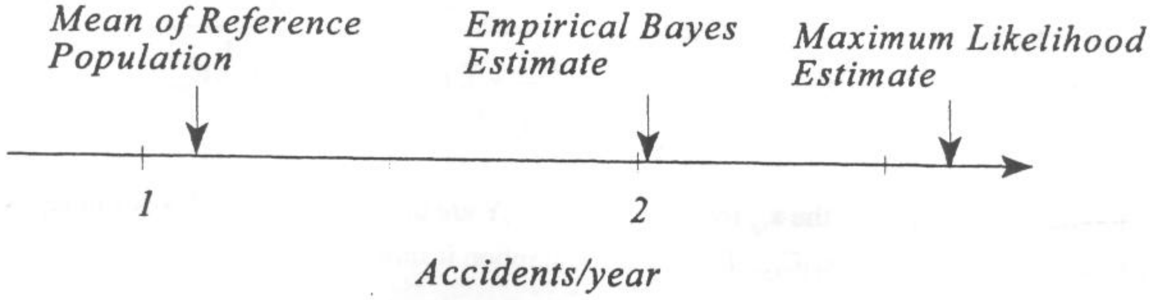


Figure 12.5. Three estimates.

Equations 12.14 are also pleasing in form and content. The parameter 'b' performs the role of an accident count; the parameter $a_i = b/E\{\kappa_{i,y}\}$ plays the role of years. In the numerical example it is as if information about the reference population amounts to 5.571 accidents in 5.571/1.115=5 years. Naturally, as the sum or real accident counts and years increases, the influence of $\hat{b}$ on the estimate of $\kappa_{i,1}$ diminishes. In the limit, the numerator of the first Equation 12.14 coincides with the maximum likelihood estimate in Equation 12.11. In addition, when the distribution of κ's in the imagined reference population is such that $\sigma\{\kappa\}$ is large compared with $E\{\kappa\}$ the EB estimate should again converge toward the maximum likelihood estimate. In this circumstance, the b in the numerator is small when compared with 1 (because $b=[E\{\kappa\}/\sigma\{\kappa\}]^2$). Also $a_i (=b/E\{\kappa_{i,1}\})$ in the denominator must be a small number when compared with 1. The smaller these two numbers are, the lesser is their influence on the estimator in Equation 12.14.

Thus, there is a pleasing internal consistency to the succession of estimators. The EB estimator (Equation 12.14) converges onto the maximum likelihood estimator (Equation 12.11) when either the information contained in the accident counts is large or when the information contained in the reference population is small. In turn, the maximum likelihood estimator converges to the ordinary estimator (the ratio of accident count and years) when the κ of the entity does not change from year to year.

There is an additional important aspect to the EB estimators in Equations 12.14 and 12.15 that is not present in the maximum likelihood estimator of Equation 12.11. Within the EB framework we can estimate the variance of the $\hat{\kappa}$'s. This ability will prove crucial when we return to the main task, that of estimating the safety effect of intervention and the precision of these estimates.

In summary, the EB estimators in Equations 12.14 and 12.15 are an attractive solution to the second of the three problems to be addressed in Part III. The first problem was that of assuming that the count of accidents alone is a sensible estimate of κ. This incorrect assumption was shown to be at the root of the regression-to-mean bias. The EB approach was the suggested response. It has been introduced in Chapter 11 and implemented again in the present chapter. The second problem was the habit of using a short 'before' period because estimation was predicated on the assumption that the κ of an entity does not change in time. From here on it is not necessary to assume so. In consequence, the concept of a 'before' period has to be revised. We now recognize that the κ of each entity changes over time. As long as we know what aspects of an entity change over time and how this change influences its κ, the entire accident history of this entity may serve for estimation. The concept of a fixed duration 'before' period as inherited from statistical experiments is thereby rendered inoperative. The duration of the 'before' period is now limited only by the earliest time for which the accident count is available, by the period to which the multivariate model applies, or by the time since when the entity exists in its present form and by the statistical judgement of what effort is worthwhile for the improvement of prediction.

The third task of Part III was to address the issue of prediction; how to make projections about what would have been the κ of a treated entity in the 'after treatment' period, had it not been treated. The estimates discussed until now are to be the basis of this prediction. How to predict is the subject of the next section.

Derivations.

Equation 12.11

The starting points are the assumption that accident counts are Poisson distributed (see Equation 12.3) and the definitional relationship $\kappa_{i,y}=\kappa_{i,1}C_{i,y}$ (see Equation 12.5). With these, as for Equation 12.8 (see Section 12.3, derivations) the probability of observing accidents counts $K_{i,1}$, $K_{i,2}$, . . . , $K_{i,Y}$ accidents in years 1, 2, . . . Y when $\kappa_{i,1}$ is given can be written as:

$$P(Accident\ Counts\ \{K_{i,y}\}|\ \kappa_{i,1}, model\ parameters,\ covariates) =$$

$$= [\prod_{y=1}^{Y} \frac{(C_{i,y})^{K_{i,y}}}{K_{i,y}!}](\kappa_{i,1}^{\sum_{y=1}^{Y} K_{i,y}} e^{-\kappa_{i,1}\sum_{y=1}^{Y} C_{i,y}}) = \mathcal{L}(\kappa_{i,1}) \qquad \text{. . . 12.8a}$$

This is the likelihood function of $\kappa_{i,1}$ when model parameters are replaced by their estimates and the covariates by their values. We wish to find that value of $\kappa_{i,1}$ at which $\mathcal{L}(\kappa_{i,1})$ is largest. Since $\mathcal{L}(\kappa_{i,1})$ is largest where $\ln\mathcal{L}(\kappa_{i,1})$ is largest, it is more convenient to examine $\ln\mathcal{L}(\kappa_{i,1})$. Thus, we need to find where the derivative $d[\ln\mathcal{L}(\kappa_{i,1})]/d\kappa_{i,1} = 0$.

$$\ln\mathcal{L}(\kappa_{i,1}) = \ln\left[\prod_{y=1}^{Y}\frac{C_{i,y}^{K_{i,y}}}{K_{i,y}!}\right] + \sum_{y=1}^{Y} K_{i,y}\ln\kappa_{i,1} - \kappa_{i,1}\sum_{y=1}^{Y} C_{i,y}$$

$$\frac{d[\ln\mathcal{L}(\kappa_{i,1})]}{d\kappa_{i,1}} = \frac{\sum_{y=1}^{Y} K_{i,y}}{\kappa_{i,y}} - \sum_{y=1}^{Y} C_{i,y}$$

If $\hat{\kappa}_{i,1}$ is that value of $\kappa_{i,1}$ at which the derivative equals 0,

$$\hat{\kappa}_{i,1} = \frac{\sum_{y=1}^{Y} K_{i,y}}{\sum_{y=1}^{Y} \hat{C}_{i,y}}$$

This is Equation 12.11.

Equation 12.13.

The starting point is Bayes Theorem, which in the present case can be written as:

$$f(\kappa_{i,1}|K_{i,1},\ldots,K_{i,Y}) = (c_1)P(K_{i,1},\ldots,K_{i,Y}|\kappa_{i,1})f(\kappa_{i,1})$$

The constant c_1 is chosen to make the integral of the conditional probability density over all non-negative values of $\kappa_{i,1}$ to be 1. The $P(K_{i,1}\ldots K_{i,y}|\kappa_{i,1})$ on the right-hand side is given by Equation 12.8a and the $f(\kappa_{i,1})$ by Equation 12.6. Upon substitution we obtain:

$$f(\kappa_{i,1}|K_{i,1},\ldots,K_{i,Y}) = (c_1)\left[\prod_{y=1}^{Y}\frac{C_{i,y}^{K_{i,y}}}{K_{i,y}!}\right]\left(\kappa_{i,1}^{\sum_{y=1}^{Y} K_{i,y}} e^{-\kappa_{i,1}\sum_{y=1}^{Y} C_{i,y}}\right)\left(\frac{a_i^b \kappa_{i,1}^{b-1} e^{-a_i\kappa_{i,1}}}{\Gamma(b)}\right)$$

We now lump c_1 with the product in the square brackets and with $a_i^b/\Gamma(b)$ into a new constant c_2. Note that c_2 that does not depend on $\kappa_{i,1}$. Thus,

$$f(\kappa_{i,1}|K_{i,1},K_{i,2},\ldots,K_{i,Y}) = (c_2)\,\kappa_{i,1}^{b+\sum_{y=1}^{Y} K_{i,y}-1} e^{-\kappa_{i,1}(a_i+\sum_{y=1}^{Y} C_{i,y})}$$

This is again a convenient result. The form of the right-hand side is the same as that in Equation 12.6. That is, $\kappa_{i,1}$ raised to some power and this expression is multiplied by 'e' raised to a power that is negative and proportional to $\kappa_{i,1}$. This means that just as $f(\kappa_{i,y})$ is a Gamma probability density function, so is $f(\kappa_{i,1}|K_{i,1}, K_{i,2}, \ldots, K_{i,Y})$. The only difference is that parameter 'a_i' of Equation 12.6 is now replaced by $a_i+\Sigma C_{i,y}$ and the parameter 'b' by $b+\Sigma K_{i,y}$.

The value of c_2 must be such that the integral over $\kappa_{i,1}\geq 0$ is 1. From here we find that

$$c_2 = \frac{(a_i+\sum_{y=1}^{Y} C_{i,y})^{b+\sum_{y=1}^{Y} K_{i,y}}}{\Gamma(b+\sum_{y=1}^{Y} K_{i,y})}$$

Therefore, the final form is:

$$f(\kappa_{i,1}|K_{i,1},K_{i,2},\ldots,K_{i,Y}) = \frac{(a_i+\sum_{y=1}^{Y} C_{i,y})^{b+\sum_{y=1}^{Y} K_{i,y}} \kappa_{i,1}^{b+\sum_{y=1}^{Y} K_{i,y}-1} e^{-\kappa_{i,1}(a_i+\sum_{y=1}^{Y} C_{i,y})}}{\Gamma(b+\sum_{y=1}^{Y} K_{i,y})}$$

Equation 12.14.

In Equation 12.6, we had $E\{\kappa_{i,1}\}= b/a_i$ and $VAR\{\kappa_{i,y}\}=b/a_i^2$. Since in Equation 12.13 parameter 'a_i' of Equation 12.6 is replaced by $a_i+\Sigma C_{i,y}$ and the parameter 'b' by $b+\Sigma K_{i,y}$ we can write,

$$E\{\kappa_{i,1}|K_{i,1},K_{i,2},\ldots,K_{i,Y}\} = \frac{b+\sum_{y=1}^{Y} K_{i,y}}{a_i+\sum_{y=1}^{Y} C_{i,y}}$$

$$VAR\{\kappa_{i,1}|K_{i,1},K_{i,2},\ldots,K_{i,Y}\} = \frac{b+\sum_{y=1}^{Y} K_{i,y}}{(a_i+\sum_{v=1}^{Y} C_{i,y})^2}$$

In Equation 12.14 parameter estimates replace their true values.

12.6 How to Predict the $\kappa_{i,Y+1}, \ldots, \kappa_{i,Y+Z}$

To learn about the safety effect of some treatment we need to predict what would have been the expected accident frequencies $\kappa_{Y+1}, \ldots, \kappa_{Y+Z}$ in the 'after' years had the treatment not been applied. Since the beginning of this chapter we have been preparing the ground for a coherent method of prediction; a method that is not restricted to the use a short 'before' period and does not use a single κ for the entire 'before' period; a method that makes efficient use of all available accident counts and is not subject to the RTM bias; a method that does not rely solely on a comparison group to convert the estimates of the 'before' period into predictions for the 'after' period.

Toward this end, the aim of predicting better, I have introduced the instrument of a multivariate model. As shown in Figure 12.1, the multivariate model is estimated from data spanning the 'before' and 'after' periods on entities representing the imagined reference populations. The uses of the multivariate model were said to be three. First, in the estimation of the moments $E\{\kappa_{i,y}\}$ and $VAR\{\kappa_{i,y}\}$ of the imagined reference population for an entity 'i'. Second, to estimate the $\kappa_{i,1}$ (or even the entire time series $\kappa_{i,1}, \kappa_{i,2}, \ldots, \kappa_{i,Y}$ for the Y 'before' years) for a treated entity 'i' that recorded $K_{i,1}, K_{i,2}, \ldots, K_{i,Y}$ accidents in the 'before' years. For this we require an assumption about how the κ's change from year to year. This too is based on the multivariate model. Finally, the multivariate model is to be used to predict the $\kappa_{i,Y+1}, \kappa_{i,Y+2}, \ldots, \kappa_{i,Y+Z}$ with the estimates $\kappa_{i,1}, \kappa_{i,2}, \ldots, \kappa_{i,Y}$ serving as the launching pad.

Now that all the pieces are in place, the final task of prediction is a bit anticlimactic in its simplicity. For a treated entity 'i' we have estimated $\kappa_{i,1}$ and its variance by Equations 12.14. Using the estimated parameters of the multivariate model and the covariate values of treated entity 'i' in the 'after' years y=Y+1, Y+2, ..., Y+Z, we compute the estimates $\hat{C}_{i,Y+1}, \hat{C}_{i,Y+2}, \ldots, \hat{C}_{i,Y+Z}$. Then, just as in Equation 12.15, for any y>Y we predict:

$$\hat{\kappa}_{i,y} = \hat{C}_{i,y}\hat{\kappa}_{i,1}$$
$$V\hat{A}R\{\hat{\kappa}_{i,y}\} = \hat{C}_{i,y}^2\, V\hat{A}R\{\hat{\kappa}_{i,1}\} \qquad \ldots 12.16$$

The first equation of the pair is a restatement of the assumption behind the rule in Equation 12.5. It reflects the belief that under the influence of the change in the covariate values of entity 'i' and the change in the factors represented by the parameters α_y of the model equation, the κ of entity 'i' is changing in the same manner as would the mean of its imagined reference population change, under the same influences. At this point an illustration may be useful.

Numerical Example 12.4. Predicting the $\kappa_{i,y}$ for a road section.

In Numerical Examples 12.2 and 12.3 I introduced a 1.6 mile long, rural, two-lane, road section 'i' in New York State. First its $\hat{C}$'s for an eight-year long 'before' period were calculated and then, based on eight accident counts, its κ's in years 1 to 8 were estimated. Consider now that this road section has been treated at the end of year 8. The task now is to predict what would have been its κ's for years 9, 10, . . ., 13, had it not been treated.

To predict, parameters of the multivariate model for the 'before' and 'after' years are needed. These were estimated using data of a sample of rural two-lane roads in New York State representing the imagined reference population, and used in the two preceding numerical examples. The first row of Table 12.6 is for the first year of the 'before' period as copied from Table 12.5. This row represents the first of the eight 'before' years. The remaining rows of Table 12.6 are for the five 'after' years (y=9 to 13). The model equation was $E\{\kappa_{i,y}\}=d_i\,\alpha_y\,F_{i,y}^{\beta}$. Estimates of model parameters α_y are listed in column 2 of Table 12.5. The additional two model parameter estimates were $\hat{\beta}$=0.776 and $\hat{b}$=5 .571. The other entries to this table will be explained below.

Table 12.6. Computations

1	2	3	4	5	6	7
y	$\hat{\alpha}_y$	$AADT_y$	$\hat{E}\{\kappa_{i,y}\}$	$\hat{C}_{i,y}$	$\hat{\kappa}_{i,y}$	$\hat{\sigma}\{\hat{\kappa}_{i,y}\}$
1	0.0028	1199	1.115	1.000	2.02	0.41
9	0.0026	1049	0.919	0.824	1.66	0.34
10	0.0027	1083	0.981	0.880	1.78	0.36
11	0.0024	1104	0.872	0.783	1.58	0.32
12	0.0025	1140	0.958	0.859	1.73	0.35
13	0.0024	1228	0.964	0.865	1.75	0.35

Numerical Example 12.4. (continued).

Using the model equation ($E\{\kappa_{i,y}\}=d_i\,\alpha_y\,F_{i,y}^{\beta}$) one can now compute estimates of the mean accident frequency of the imagined reference population of this road section. Thus, e.g., in year 9 the estimated mean is $1.6\times0.002601\times1049^{0.776}=0.919$ accidents/year. In this computation use is made of the AADT in column 3. These give the traffic flows estimated to have prevailed on the specific road section the κ's of which we eventually want to predict. Thus, the estimates of the means listed in column 4 of Table 12.6 are for imagined reference populations **exactly matched on traits** with the road section for which the prediction is needed.

Note that $\hat{E}\{\kappa_{i,1}\}=1.115$. Therefore, according to the rule in Equation 12.5, $\hat{C}_{i,9}=0.919/1.115=0.824$, $\hat{C}_{i,10}=0.981/1.115=0.880$ etc. These values are listed in column 5 of Table 12.6.

Up to this point statements have only been made about the estimated changes of the mean for the imagined reference population. It is now that we come to predicting. As has been shown repeatedly, all methods of prediction are based on some assumptions. When predicting by the Naive method one assumes that the mean of the 'before' period would have continued into the 'after' period. When prediction is by the Comparison-Group method one assumes that the mean of the treated group would have changed from 'before' to 'after' in the same ratio as the mean of the comparison group has. The same need to make assumptions about what would have happened applies here. I will try to clarify what exactly is being assumed.

Numerical Example 12.4 . (continued) - Assumption and prediction.

In Table 12.6 we see that the mean of the imagined reference population for road section 'i' has changed from year 1 to year 9 by a factor $\hat{C}_{i,9}=0.824$. This is partly because AADT has changed from 1199 to 1049, and partly because the change in all the other influences, as reflected in a change from $\hat{\alpha}_1=0.002844$ in year 1 to $\hat{\alpha}_9=0.002601$ in year 9 Road section 'i' has undergone the same changes. The change of the AADT is exactly that in the reference population for which $\hat{C}_9$ was computed. In addition, because our road section is not distinguishable from members of the imagined reference population, we are assuming that its α has changed in the same way as that of the reference population. It is for these reasons that the ratio $\kappa_{i,9}/\kappa_{i,1}$ is assumed to be the same as the ratio $E\{\kappa_{i,9}\}/E\{\kappa_{i,1}\}=C_{i,9}$. This then is the assumption on which the prediction and Equations 12.16 are based.

At this point I reach back to Numerical Example 12.3 where we estimated that road sections with these traits which in the eight 'before' years has recorded 19 accidents, are estimated to have $\hat{\kappa}_1$=2.02 accidents/year. This value is also shown in the first row of Table 12.6. If so, by Equation 12.6, $\hat{\kappa}_{i,9}=\hat{C}_{i,9}\hat{\kappa}_{i,1}$= 0.824×2.02=1.66 accidents/year. Predictions for years 10 to 13 are computed similarly and are listed in column 6 of Table 12.6.

It may be useful now to draw some parallels. In Chapter 5 I have said that to predict well one must find answers to the following questions: **a.** How to account for the change in time of those causal factors that affect safety, are recognized, measured, and the influence of which is known and; **b.** How to account for the change in time of all the remaining causal factors which are either unrecognized, not measured, or the influence of which on safety is not sufficiently understood to be quantified. In Chapter 8 I have resolved to account for the recognized-measured-understood-factors explicitly, that is by quantitative modeling. Thus, e.g., I suggested to account for changes in traffic flow by using a 'safety performance function' which links traffic flow to expected accidents. For the remaining causal factors, the unrecognized-not measured-not quantified group, I suggested to account implicitly, by means such as the use of a comparison group or by extrapolation of time trends. This was the subject of Chapter 9.

Unfortunately, this two pronged approach led to a practical difficulty (see Section 9.6). If a quantitative model is used first to account for changes in, say, the traffic flow, one may not then simply go on to use a comparison group and thereby account for the change in the remaining causal factors. Inasmuch as the comparison group was also subject to changes in traffic, steps must be taken to avoid 'double accounting'. In concept this is easy to do, but in practice the traffic data for all entities in the comparison group are unlikely to be available.

The method of prediction suggested here is entirely consistent with this two-pronged strategy. It is pleasing, that it turned out so simple and problem free. The factors $C_{i,y}$ at once account for both kinds of causal factors. The recognized-measured-understood factors form that part of the model equation which consists of all the covariates and their parameters. The change in the unrecognized-not measured-not understood factors is represented by the parameters α_y of the model equation. Both parts of the model equation feature in the estimation of the $C_{i,y}$'s. In this manner, the multivariate model serves the dual purpose of being at once a 'safety performance function' and also a comparison group. In all this, there is no threat of double accounting. The last topic to be discussed here is the precision of the predictions.

Numerical example 12.4. (continued) - Prediction precision.

Use here $\hat{\sigma}\{\hat{\kappa}_{i,y}\} = \hat{C}_{i,y} \times \hat{\sigma}\{\hat{\kappa}_{,1}\}$ from the second of the pair in Equations 12.16. From Numerical Example 12.3 and also the first row in Table 12.6 we have $\hat{\sigma}\{\hat{\kappa}_{i,1}\}=0.41$ accidents/year. Thus, e.g., $\hat{\sigma}\{\hat{\kappa}_{i,9}\}=0.824\times0.41= 0.34$. The estimated standard deviation of the predictions is shown in the cross-hatched part of column 7 in Table 12.6.

Normally, the further one predicts into the future, the less certain is the prediction. Therefore, it may look peculiar that the precision of the prediction in Table 12.6 does not deteriorate the further one goes into the unobserved future of this road section. The explanation consists of two parts. First, not all aspects of the years 9 to 13 are 'unobserved'. Specifically, we have AADTs for these 'future' years and we have estimates of the α's telling how the sundry factors influenced the safety of the reference population. Second, and more importantly, the standard deviations are computed **as if the assumptions were true**. That is, as if the model equation correctly represented the influence of all important variables, and as if the parameter estimates were close to true, and as if the assumption on which the prediction is based was correct, and as if this road section in fact had no features or influences to distinguish it from the reference population, etc. In short, the estimated standard deviation is, as always, a 'conditional' one. In the final account, a realistic estimate of the precision of predictions can only be obtained by empirical means. No doubt that when this is done, the actual standard deviations of predictions will be found larger than those computed by Equation 12.6.

The aim of Chapter 12 was to provide a more coherent approach to the conduct of observational studies than the patchwork of adaptations of which Part II is made up. This task is now complete. It remains to describe how the suggested approach has been used in a specific case.

12.7 THE SAFETY EFFECT OF ROAD RESURFACING IN NEW YORK STATE

Layer upon layer, the last six sections built up a coherent approach for the conduct of observational Before-After studies. Finally, in the last section, the keystone has been put in place. It is now clear how to predict what would have been the safety of an entity in the 'after' period had the treatment not been applied. This prediction can then be compared with the estimate of what was the safety of the entity with the treatment in place. By comparing 'what was' and 'what would have been', the effect of the treatment on the safety of the treated entity is established. When this is done for each treated entity of a group, one can make statements about the average effect of the treatment on this group and perhaps also about the variability of the effect within the group. To the extent that the treated group is similar to some entities yet untreated, one can also anticipate what might be the safety effect of the treatment, if implemented. To round off this chapter I will recount the story of how the suggested approach has been implemented in a specific circumstance. A full description may be found in Hauer et al., (1995b).

In the early eighties, two kinds of pavement resurfacing projects were undertaken in New York State: 'Fast Track' (FT) projects involving only resurfacing, and 'Reconditioning and Preservation' (R&P) projects where roadside and roadway safety improvements have been incorporated with resurfacing. The question was whether following resurfacing the FT projects (82 road sections, 226.7 miles) performed worse, from a safety viewpoint, than the R&P projects (55 road sections, 137.2 miles). The effect of these treatments on intersection accidents, non-intersection accidents, and on fixed-object accidents has been examined. In this section I will describe only the estimation of the effect on non-intersection accidents.

The data used pertained to rural, two-lane, undivided, free-access road sections. For each road section the following information has been assembled:

1. The length of the section and the number of intersections in it.
2. Traffic counts for the 13 years 1975-1987 factored to represent the AADT in the year of the count.
3. The count of accidents for each month of the 13 years from 1975 to 1987.
4. If the road section was resurfaced, the month and year in which construction started and ended. (Mostly in the 1981 and 1982 construction seasons)

In addition to the treated road sections, the same data was available for a sample of 525 reference road sections (2193.2 miles). These are representing the rural two-lane road system in New York State. A model of the form $E\{\kappa_{i,y}\}=d_i\alpha_y F_{i,y}^{\beta}$ has been fitted to this data. Its parameter estimates have already been given in Numerical Examples 12.2 and 12.4. To describe how the safety effect of resurfacing has been estimated I will make use the results of these earlier numerical examples. The entity used in these examples was one of the treated road sections. Its pavement has been resurfaced during April to November of 1982. The fact that resurfacing took place during the year and not at its end will require some slight modification of earlier results. This will add a touch of reality.

On this road section, for the 'before resurfacing period', we had seven full years (1975-1981) and the first three months (in 1982). Thus, while the number of before periods is Y=8, as earlier, the fact that the eight's period was only 3-months long requires an adjustment. In Table 12.6, $\hat{C}_{i,8}=0.849$. Recall that the C's were akin to counters of years when the yearly accident counts K were used for estimation. To account for the fact that the accident count in 1982 was for only a quarter of a year, we use here $\hat{C}_{i,8}$=0.849/4=0.212. The accident count should also be adjusted to be that for January to March 1982. But, since there were no accidents on this road section in 1982, no adjustment is needed. Therefore, using the information in Table 12.5, the sum of C's here is 7.173-0.849+0.212= 6.536 (and not 7.173 as in Numerical Example 12.3). With this modification, by the procedure illustrated in Numerical Example 12.3, $\hat{\kappa}_{i,1}$=(5.571+19)/ (5.571/1.115+6.536)=2.13 accidents/year in 1975 (instead of 2.02). Also, $\hat{\sigma}\{\hat{\kappa}_{i,1}\}$= [2.13/(5.571/1.115+6.536)]$^{0.5}$=0.43 accidents/year.

We can now predict what the expected number of accidents on this site would have been had no resurfacing taken place. Thus, e.g., for the first full year after resurfacing $\hat{\kappa}_{i,1983}$= 0.824×2.13=1.76 accidents/year with a standard deviation of 0.824×0.43=0.35 accidents/year. The next step is obvious. In 1983, after the site has been resurfaced, it recorded three non-intersection accidents. Without resurfacing we would have expected to have 1.76 such accidents. Thus, for this site and year, there were 1.24 more non-intersection accidents than expected. Since changes in traffic and other factors were accounted for, the noted difference is attributable to resurfacing. Of course, one will not form an opinion about the safety effect of resurfacing based on one site and one year. The effect will be added up for all sites and examined for all years.

So far in this chapter, analysis and estimation were based on a division of time into years. For the 'resurfacing in New York State' this choice was dictated by the fact that data about the AADT was available only on a yearly basis. However, since the effect of resurfacing is likely to change with time, it seemed important to examine the effect on a monthly, rather than yearly basis. Indices of monthly variation in non-intersection accidents on rural two-lane roads in New York State are given in Table 12.7. Thus, e.g., Januaries will have, on the average, 1.25 times the yearly accidents/12.

Table 12.7. Monthly variation indices for non-intersection accidents in New York State.

Month											
1	2	3	4	5	6	7	8	9	10	11	12
1.25	0.97	0.97	0.79	0.84	0.89	0.96	0.98	0.85	0.95	1.15	1.41

To illustrate the use of these indices for creating estimates of monthly projections, consider again the road section for which we projected 1.76 accidents. Therefore, e.g., for May of 1983, multiply 1.76 by 0.84/12. Thus, for May 1983 we project for this road section 0.123±0.025 accidents (per month). Similar monthly projections can be made for every month of 1983. For 1984 we project $\hat{\kappa}_{i,1984}$= 0.880×2.13=1.87 accident/year and therefore, say, for January of 1984 we project 1.87×1.25/12=0.196 accidents/month. The monthly projections both during and after construction are shown in Figure 12.6. We will be interested mainly in the 'after resurfacing' months. The first month after resurfacing will have the counter index 1. For this road section there are 63 months after resurfacing.

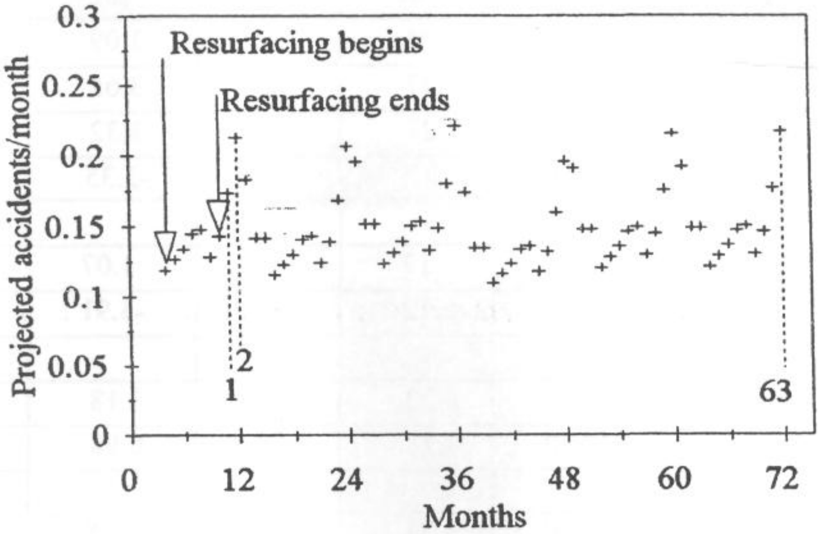


Figure 12.6. Monthly projections of the expected number of non-intersection accidents.

Similar monthly projections were prepared for each treated road section. Adding the projections for the first month after resurfacing over all 82 road sections in the 'Fast Track' (FT) projects group, we obtained 20.91 non-intersection accidents. That is, 20.91 accidents would be expected in the first month after the resurfacing had no resurfacing taken place. This is the first entry of column 2 in Table 12.8. The second entry in column 2 is the projected number of accidents in the second month after resurfacing (23.35) and so on. In column 3 I show the cumulative sum of the projections in column 2. The reason for this cumulation will become apparent shortly. In column 4 I list the number of accidents that have occurred on the 82 FT road sections. Thus, there was a total of 24 accidents in the first month after resurfacing, 27 accidents in the second month and so on. Both the 20.91 and the 24 in the first row are the sums over all the FT road sections. This is the number in column 5. However, in the lower reaches of the table, the sums are over fewer road sections. This reflects the fact that the later a road section has been resurfaced, the fewer 'after months' will there be since we have no data after the end of 1987. Thus, e.g., the projections in Figure 12.6 for the road section 'i' of the earlier numerical examples go only up to 63 months after resurfacing. Only 40 of the 82 FT road sections had an available data history longer than 74 months after resurfacing.

Table 12.8. Summary calculations.

Months After Resurfacing (1)	Projected Number of Accidents (2)	Cumulative Projected Accidents (3)	Monthly Accident Count (4)	Number of Road Sections (5)	Excess Accidents (6)	Cumulative Excess (7)
1	20.91	20.91	24	82	3.09	3.09
2	23.35	44.26	27	82	3.65	6.74
3	23.68	67.94	27	82	3.32	10.06
4	20.35	88.29	17	82	-3.35	6.71
...						
30	18.07	605.37	17	82	-1.07	124.63
31	17.91	623.27	12	82	-5.91	118.73
...						
41	18.82	829.78	22	82	3.18	138.22
42	17.02	846.80	18	82	0.98	139.20
...						
60	21.08	1203.41	21	82	-0.08	133.59
61	23.57	1226.98	31	82	7.43	141.02
...						
73	16.24	1438.00	12	50	-4.24	106.00
74	16.97	1454.97	20	40	3.03	109.03

The difference between the estimate of 'what was' in the first month after resurfacing (24 accidents), and the projection of 'what would have been' (20.91 accidents), is the estimate of the 'excess number of accidents'. This is listed in column 6. In the terminology of Chapter 6 it is $-\hat{\delta}$. The last column lists the cumulative excess. Thus, at the end of the third month after resurfacing, the cumulative excess is estimated to be 10.06 accidents. The data of Table 12.8 are shown in Figure 12.7. The crosses belong to the left scale and show the accumulation of the excess of non-intersection accidents with time after resurfacing (the last column in Table 12.8). The solid line in Figure 12.7 belongs to the right scale and shows the accumulation of the number of accidents expected without resurfacing.

The orderliness of the results, when presented in this cumulative form, is quite remarkable. For the first 30 months or so, there is a monthly excess averaging 4.15 non-intersection accidents. The sample standard deviation of this average, as calculated from the thirty monthly values, is 0.93. So, even from the purely statistical point of view, the excess must be thought real (the hypothesis that there was no increase in non-intersection accidents is clearly rejected). Over the first 30 months the excess accumulates to about 125 non-intersection accidents with a standard deviation of 28. Thus, we estimate that had no resurfacing taken place, and if the pre-resurfacing pavement conditions continued to prevail, 125±28 fewer non-intersection accidents would have been recorded within 30 months. Over the same period about 605 non-intersection accidents would be expected without resurfacing. Thus, the increase is of about 21%±5% (125/605=0.21, 28/605=0.05).

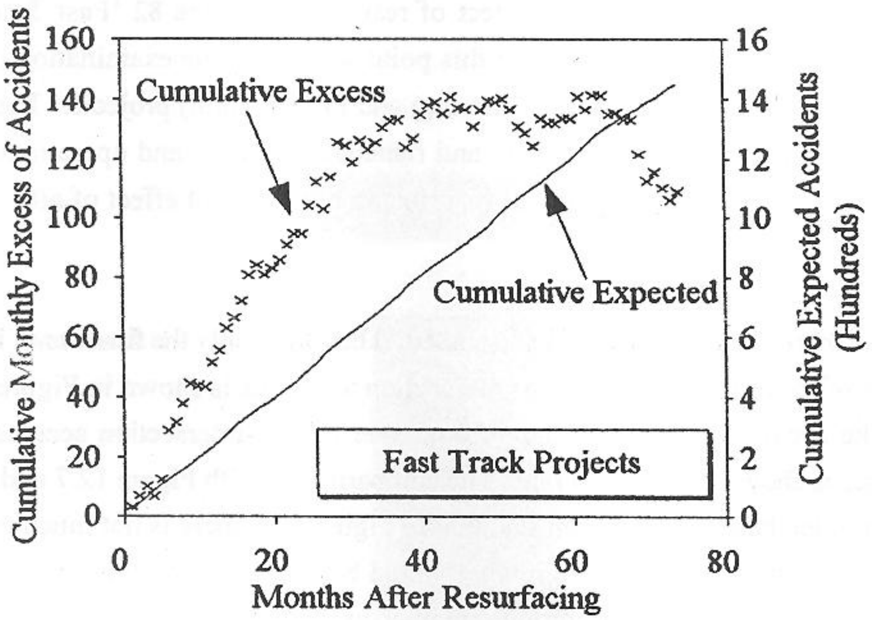


Figure 12.7. The safety effect of resurfacing on non-intersection accidents in 82 Fast Track projects.

After the first 30 months there is a 10 months transition during which the monthly accident excess gradually diminishes. The detrimental effect of resurfacing seems to vanish after about 40 months where the cumulative excess graph begins to be roughly horizontal. Over these 40 months an excess of some 135 non-intersection accidents has accumulated, with a standard deviation of 33. Without resurfacing one would have expected to accumulate by that time 810 non-intersection accidents. Thus, over a period of 40 months, the resurfacing resulted in an increase of about 17% in non-intersection accidents in the 'Fast Track' projects.

After 40 months there is a plateau that lasts till about 63 months after resurfacing. During this period the average monthly excess is of 0.28 non-intersection accidents. The standard deviation of this average is 0.79 accidents. Thus, it appears that on the plateau the number of accidents is approximately what would have been expected without resurfacing but with pavement conditions that prevailed before resurfacing. Following the plateau there is a gradual decline. That is, beginning with month 64 after resurfacing, fewer non-intersection accidents occur than one should expect if the pre-resurfacing pavement conditions continued to prevail. The number of sites having such a long post-resurfacing history is small. Therefore, one cannot say whether the noted decline is real. However, similar declines were noted for the other accident types as well. One may speculate that after more than five years of service, the pavement condition is on the average worse than what it was in the before-resurfacing period. Just as a new pavement was seen to generate an excess of non-intersection accidents, it should not be surprising that old pavements seem to have the opposite effect.

So far the focus was on the safety effect of resurfacing on the 82 'Fast Track' projects involving only resurfacing and re-striping. At this point we turn to the examination of the safety effect of resurfacing on the 55 R&P (Reconditioning and Preservation) projects. These typically entailed also limited pavement reconstruction, and remedies to safety and operational problems. Thus, the effect to be estimated is not only of resurfacing but the joint effect of all modifications implemented.

The method here mimics that already discussed. Therefore, only the final result is of interest. The accumulation of the monthly excess non-intersection accidents is shown in Figure 12.8 by the 'cross' signs. The accumulation of the monthly number of non-intersection accidents expected without resurfacing is shown by the solid line. The comparison is with Figure 12.7 and the contrast is stark. Without much hesitation one can say, that in Figure 12.8 there is not much change in the number of non-intersection accidents from what would be expected had these projects not been implemented and had the pavement condition remained as in the pre-improvement period.

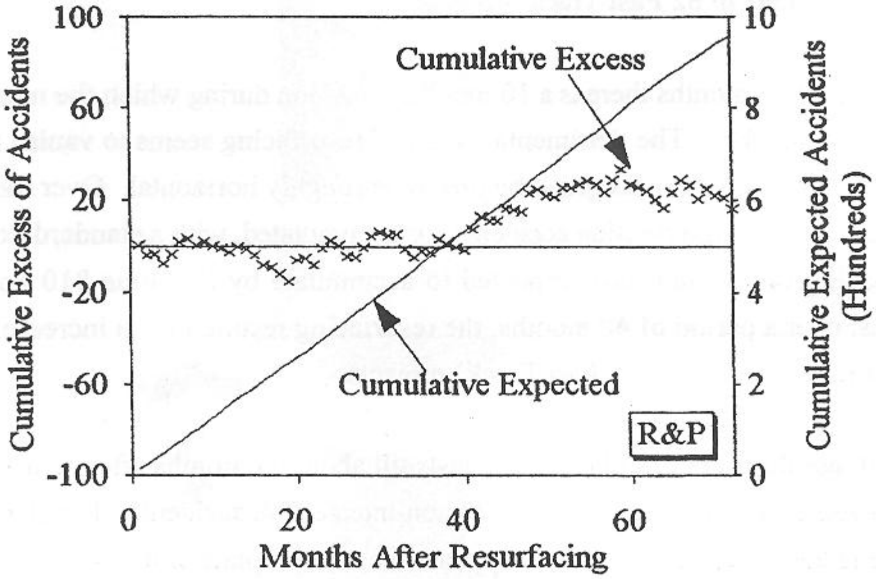


Figure 12.8. The safety effect of resurfacing and other improvements on non-intersection accidents in 55 R&P projects.

It seems clear that the kind of 'resurfacing' which went with FT projects affected safety differently from the kind of 'resurfacing' associated with R&P projects. This leads to the conclusion that 'resurfacing' as referred to in the professional literature may cover a heterogeneous set of activities.

There is a methodological deficiency to the study, which became apparent after the results were obtained, but should have been anticipated. From the outset it was sensible to expect that the effect of resurfacing on safety changes as the pavement ages. This is why the examination was month-by-month. Since the road sections were all resurfaced in the 1981-1982 construction season, that is, within one year of each other, one may think that they were all in need of repair before resurfacing and in reasonably good shape 5-7 years earlier. If so, there is a systematic factor which our analysis neglected. To predict what would have happened had these roads not been resurfaced, one should have accounted for the change in safety as the pavement would have deteriorated further. This factor is not accounted for by the α's of the model equation nor by the covariates used. The net effect of this oversight in method is, that the predictions have been produced as if a constant average pavement condition prevailed during the entire before-resurfacing period and the same would have continued had the roads not been resurfaced. The incorporation of a pavement-age covariate in the model equation, would have been an appropriate remedy. I suspect, however, that such data was not readily available.

The purpose of this section was to tie things together; to show how the method of prediction developed in this chapter can be used, and how it has been applied in a specific circumstance. In this case, it seems to have been quite successful in extracting clear conclusions from the kind of data that often leads to ambiguity. One could, no doubt, interpret the same data by a traditional C-G approach in which the reference road sections would be the comparison group. However, doing so would not account for the possible RTM bias arising out of the possibility that the accident history of a road section influenced the decision to resurface it. Nor would it be possible to account cleanly and convincingly for changes in traffic flow and separately for the yearly effect of all other causal factors. Most importantly, one could not obtain by the C-G method month-by-month estimates such as those shown in Figures 12.7 and 12.8. The available reference group, although respectably large, cannot furnish a sufficient number of accidents to estimate statistically reliable monthly comparison ratios. Nor would it have been possible to base the projection on 7-8 years of 'before' data.

12.8 CHAPTER SUMMARY

An observational study is not a statistical experiment. Therefore some habits of thought shaped by familiarity with methods for the interpretation of statistical experiments, need to be critically reviewed, perhaps jettisoned. Thus, it is not natural to think that the 'before' period has some fixed duration; nor it is not useful to assume that in the 'before' period there was a single expected accident frequency κ; neither it is attractive to use a single 'comparison ratio' to account for change from 'before' to 'after' in all the causal factors. In this chapter I suggest a coherent approach to the conduct of an observational Before-After study.

The element that ties the pieces together is a multivariate model estimated from a sample of untreated entities and represents the reference populations of the treated entities. First, the multivariate model allows one to estimate the $E\{\kappa_{i,y}\}$ and $VAR\{\kappa_{i,y}\}$which are needed to eliminate the regression-to-mean bias from estimation. Second, the multivariate model allows one to estimate the factors $C_{i,y}$ which are needed in order to use many years of accident counts in estimation. Third, the multivariate model allows one to predict what would have been the $\kappa_{i,Y+1}, \ldots, \kappa_{i,Y+Z}$ in the 'after' period taking into account the changes in measured covariates and also changes in unmeasured factors.

The suggested method is based on a small number of plausible assumptions, it is logically coherent, it is based on a few simple equations that have an intuitive interpretation and appeal, and, above all, it is practically workable. However, while the method is workable, it is not as simple as, say, the routine analysis of a 2×2 table of the C-G kind.

First, one needs to have data for a sample from the reference population. This requirement, in itself, is not much different from the need to have the requisite data for, say, a comparison group. What is new, however, is the need to fit a multivariate model to the reference group data. This task should not be performed by persons untrained in the craft. In particular, the choice of model form and the need to obtain parameter estimates for the year-to-year variation (the α_y) and 'b' require substantive expertise.

Second, as illustrated in the last section, an observational study is of the nature of a 'variation on a theme'. There is the basic sequence of actions - the fitting of a multivariate model, the estimating of the $\kappa_{i,1}$, the predicting of the $\kappa_{i,Y+1}, \ldots, \kappa_{i,Y+Z}$ - this is the 'theme'. But then, there is also the apparent need to make projections monthly, to exclude construction periods, not to forget to account for pavement aging etc. - these are the 'variations'. Statistical experiments, studies that are premeditated, designed, and involve random assignment for treatment, are perhaps reducible to routine. This, at least, is the impression left by books on the subject. Observational studies which make use of what data are possible to assemble, and do not entail random assignment to treatment, cannot be interpreted by following prescriptions, as if they were merely second class statistical experiments. The best one can do is to suggest the structure of the 'theme', the needed variations are for the expert analyst to provide.

CHAPTER 13
CLOSURE

A 'final endnote' might have been a better title for this chapter, because the subject of how to interpret observational Before-After studies has certainly not reached closure. Not much has been written on this topic, and a great deal needs to been done before a maturity of method and routine of application can be attained or claimed.

It is indeed a curious state of affairs. The overwhelming majority of accounts about the safety effect of this or that treatment published in the professional literature, are observational studies, not experiments. The data we use in research on safety are almost always what Box et al. (1978, 487-498) call 'happenstance data'. That is, there is no premeditated selection of the levels of covariates at which to observe the response variable; there is no random assignment to treatment or control. I suspect that the same is true not only of safety, but also in many applied fields of engineering and social science. Yet, compared with the acres of books on the statistical design and interpretation of experiments, books on observational studies can be counted on the fingers of one hand. If the need exists, why is it not met?

The answer, I think, has three facets. First, judging by what is usually done, the assumption must be that the tried and tested statistical methods used to interpret experiments, also apply, or are the best choice for, the interpretation of observational studies. In some cases this assumption reflects statistical naivete. One or two courses in statistics are sufficient to acquaint us with the tools of regression or hypothesis testing, but they seldom create an adequate awareness of the fact that these tools can be used only on random samples and data from experiments. In other cases, the statistically sophisticated, those who know the difference between data from experiments and data from observational studies, must believe that even happenstance data is still best interpreted by the classical methods. The edifice of classical statistical methods is indeed an impressive one. However, it does not contain any support for such a belief and plenty of caution against it. On this score then, the paucity of methods for observational studies can be explained but not justified.

The second facet is a pessimistic one. It is possible that so little has been written about how to interpret happenstance data because most believe such data to be uniterpretable. The belief must be that phenomena involving considerable randomness can be interpreted and understood only through experiments and not through observational studies. If this opinion is correct, one should not be able to learn much about the connection between smoking and lung cancer by observing the incidence of this disease in the population. Some statisticians still cling to this opinion, but it is a minority view. If so, most must think that observational studies can lead to useful conclusions. Therefore, on this score too, methods for interpreting observational studies deserve much more attention than are getting.

The third possible explanation is that the statistical interpretation of observational studies is messy, involves ambiguity, may require judgement and, in general, does not provide the intellectual pleasures of clear logic, systematic deduction and uncontrovertible proof. As a result it may be an unattractive field of endeavour for the disciplined mind that aims to share in the delights of science. To this I can only respond, weakly, that the interpretation of observational studies is both more challenging than the interpretation of experiments and has been examined very little. It should therefore be an attractive field of inquiry for the best minds.

It is my belief that observational studies can lead to correct insights about the world, in particular, about the safety effect of treatments. In the field of road safety, in which there is so little experimentation and so much rich and growing experience in implementation, the interpretation of observational studies is a key to progress. This does not mean that one should not prefer the conduct of an experiment to an observational study. It does mean, however, that the available data generated in the course of the countless instances of treatment implementation should be well exploited. It is my hope that this book provides useful guidance in this matter.

REFERENCES

Abbess, C., Jarrett, D., and Wright, C.C. (1981). Accidents at blackspots: estimating the effectiveness of remedial treatment, with special reference to the "regression-to-the-mean" effect. *Traffic Engineering and Control,* **22,** No. 10, 535-542.

Ang, A. H-S and Tang, W.H. (1975). *Probability concepts in engineering,planning and design.* John Wiley and Sons, New York.

Agran, P.F. and D.E. Dunkle (1985). A comparison of reported and unreported noncrash events. *Accident Analysis and Prevention,* **17,** No. 1, 7-13.

Andersson, K. and Nilsson G. (1981). The effect on accidents of compulsory use of running lights during daylight in Sweden. Statens väg-och trafikinstitut (VTI), Report No. 208A.

Benjamin, J.R. and Cornell, C.A. (1970). *Probability, statistics and decision for civil engineers.* McGraw-Hill, New York.

Berkson, J.,(1942). Tests of significance considered as evidence. *Journal of the American Statistical Association.* **37**, 325-335.

Bhesania, R.P. (1991). Impact of mast-mounted signal heads on accident reduction. *ITE Journal,* **61**, 10, pp.25-29.

Box, G.E., Hunter, W.G., Hunter S.J., (1978). *Statistics for experimenters.* John Wiley & Sons, New York.

Box, G.E.P. (1966). Use and abuse of regression. *Technometrics,* **8**, 625-629.

Brundell-Freij, K., and Ekman, L., (1991). Flow and safety. Paper 910006, presented at the 70th Annual Meeting of the Transportation Research Board in Washington D.C.

Clark, J.E., Maghsoodloo, S., and Brown, D.B. (1983). The public good relative to right turn on red in South Carolina and Alabama. *Transportation Research Record 926,* 24-31. Transportation Research Board, Washington, D.C.

Campbell, B.J., (1968). Seat belts and injury reduction in 1967 North Carolina automobile accidents. Highway Safety Research Center, Chapel Hill, NC.

Campbell, D.T. and Stanley, J.C. (1966). *Experimental and quasi-experimental desings for research.* Rand McNally College Publishing Company, Chicago.

Cochran, W.G. (1954). The combination of estimates from different experiments. *Biometrics,* **10,** 101-129.

Cochran, W.G. (1983). *Planning and analysis of observational studies.* Edited by Moses, L.E. and Mosteller, F., John Wiley & Sons. New York.

Comissioner of Public Works and the Environment (1991). Stop sign management program. Unpublished memorandum. City of Toronto.

Cook, T.D. and Campbell, D.T. (1979). *Quasi-experimentation, Design and analysis issues for field settings.* Rand McNally, Chicago and Houghton Mifflin, Boston.

Cooper, H.M., (1989). *Integrating research: A guide for literature reviews.* Sage, Beverly Hills, California.

Edwards, A.W. F. (1972). *Likelihood; an account of the statistical concept of likelihood and its application to scientific inference.* Cambridge University Press.

Efron B. and Morris C. (1977). Stein's paradox in statistics. *Scientific American,* **236,** 119-128.

Elvik, R., Vaa, T., Østvik, E., (1989). Traffikksikkerhetshanbok. (Traffic safety handbook). Revised edition, Institute of Transport Economics, Oslo.

Elvik, R. (1994). Meta analyse av effektmalinger av trafiksikkerhetstiltak. (Meta-analysis of evaluations of the effectiveness of road safety measures.) TØI-rapport 232. Institute of Transport Economics, Oslo.

Elvik, R., (1995a). The safety value of guard rails and crash cushions: a meta analysis of evidence from evaluation studies. *Accident Analysis and Prevention.* **27,** No. 4, 523-549.

Elvik, R. (1995b). Meta-Analysis of evaluations of public lighting as accident countermeasure. *Transportation Research Record 1485,*112-123. Transportation Research Board, Washington, D.C.

Evans, L. (1991). Older driver risks to themselves and to other road users. *Transportation Research Record 1325,* 34-41. Transportation Research Board, Washington, D.C.

Evans, L. (1993). Medical accidents: no such thing? *British Medical Journal,* **307,** No. 6917, 1438-1439.

FHWA (1981). Accident reduction factors. Technology sharing report TS-81-209. U.S. Department of Transportation. Federal Highway Administration. Washington, D.C.

FHWA (1982). Synthesis of safety research related to traffic control and roadway elements. Volumes 1 and 2. Reports FHWA-TS-82-232 & 233 U.S. Department of Transportation. Federal Highway Administration, Washington D.C.

FHWA (1988). Manual on Uniform Traffic Control Devices for Streets and Highways. ANSI D6.1e-1989. Federal Highway Administration, Department of Transportation. Washington, D.C.

FHWA, (1992). Safety effectiveness of highway design features. Volumes 1 to 6. Publication No. FHWA-RD-91-044 to 049. U.S. Department of Transportation. Federal Highway Administration. Washington D.C.

Fisher, R.A. (1971). *The design of experiment.,* Hafner Publishing Company, New York. First published in 1935.

Fleiss, J.L. (1981). *Statistical methods for rates and proportions.* Second edition, John Wiley & Sons, New York.

Galton Sir, F. (1877). Typical Laws of Heredity. *Proc. Roy. Inst.,* **8,** 282-301.

Gartner, N.H., Messer, C.J., and Rathi, A.K. (eds.) (1997?). Traffic flow theory. U.S. Department of Transportation. Federal Highway Administration. Draft. (To be published as a new edition of Transportation Research Board Special Report 165).

Glass, G.V. (1976). Primary, secondary and Meta-Analysis of research. *Educ. Res.*, **5**, 3-8.

Glass, G.V., McGraw, B. and Smith, M.L. (1981). *Meta-analysis in social research.* Sage. Beverley Hills, California.

GLIM (1987) The generalized linear interactive modelling system. Numerical Algorithms Group Inc., ISBN 1-85206-038-7.

Griffin III, L.I. (1989). A systematical framework for analyzing categorical, before-and-after data. Safety Division, Texas Transportation Institute, College Station, Texas.

Griffin III, L.I. (1990 a). Estimating the effectiveness of occupant protection safety devices from state accident data. Conference on the Collection and Analysis of State Highway Data. San Diego.

Griffin III, L.I. (1990 b). Using the Before-After design with yoked comparisons to estimate the effectiveness of accident countermeasures implemented at multiple treatment locations. Texas Transportation Institute. College Station. Texas.

Hakkert, A.S., Livneh, M. and Mahalel, D. (1976). Levels of safety in accident studies - a safety index. ARRB Proceedings, **8**, Part 5, 1-6. Australian Road Research Board, Victoria.

Hauer, E. (1980a). Bias-by-selection: Overestimation of the effectiveness of safety Counter-measures caused by the process of selection for treatment. *Accident Analysis and Prevention*, **12**, No.2, 113-118.

Hauer, E. (1980b). Selection for treatment as a source of bias in before-and-after studies. *Traffic Engineering and Control*, **21**, Nos. 8/9, 419-421.

Hauer E. and Persaud B.N. (1983). A common bias in before and after accident comparisons and its elimination. *Transportation Research Record 905*, 164-174. Transportation Research Board, Washington, D.C.

Hauer, E. (1983a) Reflections on methods of statistical inference in research on the effect of safety countermeasures. *Accident Analysis and Prevention*, **15**, No. 4, 275-285.

Hauer, E., Lovell, J., and Persaud, B.N. (1984). The safety effect of conversion from two-way to four-way stop control in San Francisco. Publication 84-05, Dept. of Civil Engineering, University of Toronto, Toronto.

Hauer, E. (1986a). On the estimation of the expected number of accidents. *Accident Analysis & Prevention*, **18**, 1-12.

Hauer, E. and Lovell, J. (1986b). New directions for learning about the safety effect of measures. *Transportation Research Record 1068*, 96-102. Transportation Research Board, Washington, D.C.

Hauer, E. and Persaud, B.N. (1987). How to estimate the safety of rail-highway grade crossings and the safety effect of warning devices. *Transportation Research Record 1114*, 131-140. Transportation Research Board, Washington, D.C.

Hauer, E. (1988). A case for science-based road safety design and management. In: Stammer, R.E., (ed.), *Highway Safety: At the crossroads.* American Society of Civil Engineers, New York.

Hauer, E. and Hakkert, A.S. (1989). The extent and implications of incomplete accident reporting, *Transportation Research Record 1185*, 1-10, Transportation Research Board, Washington D.C.

Hauer,E., Ng, J.C.N and Lovell, J, (1989). Estimation of safety at signalized intersections. *Transportation Research Record 1185*, 48-61. Transportation Research Board, Washington, D.C.

Hauer, E. (1989). The multivariate method for estimation of unsafety, Department of Civil Engineering, University of Toronto, Report No. 89-09, ISBN-0-7727-7556-7.

Hauer, E. (1990). Empirical Bayes approach to the estimation of 'unsafety': the multivariate method. Report No. FHWA-RD-90-006, US. Department of Transportation. Federal Highway Administration, Washington, D.C.

Hauer, E., Ng, J.C.N and Papaioannou, P. (1991). Prediction in road safety studies. *Accident Analysis & Prevention*, **23**, No. 6, 595-608.

Hauer, E. (1991a). Should Stop Yield? Matter of method in safety research. *ITE Journal*, **61**, No. 9, .25-32.

Hauer, E. (1991b). Comparison groups in road safety studies; An analysis. *Accident Analysis & Prevention*, **23**, No. 6, 609-622.

Hauer, E. (1992a). A note on three estimators of safety effect. *Traffic Engineering and Control*, **33**, No. 6, 388-393.

Hauer, E., (1992b). Empirical Bayes approach to the estimation of unsafety: The multivariate regression approach. *Accident Analysis & Prevention*, **24**, No. 5, 456-478.

Hauer, E. (1993). Overview. In: ITE, The traffic safety toolbox; A primer on traffic safety. Institute of Transportation Engineers, Washington, D.C.

Hauer, E., D. Terry,M. S.Griffith (1995a),The effect of resurfacing on the safety of two-lane rural roads in New York State. *Transportation Research Record 1467*, 1994, 30-37. Transportation Research Board, Washington, D.C.

Hauer, E. (1995b). Exposure and accident rate. *Traffic Engineering and Control*, **36**, No. 3, 134-138.

Hauer E. and Persaud, B.N. (1996). Safety analysis of roadway geometry and ancillary features. Transportation Association of Canada. Ottawa.

Hauer, E. (1996a). Detection of safety deterioration in a series of accident counts. *Transportation Research Record 1542*. 38-43. Transportation Research Board. Washington, D.C.

Hauer, E. (1996b). Statistical test of a difference between expected accident frequencies. *Transportation Research Record 1542*. 24-29. Transportation Research Board. Washington, D.C.

Hauer, E. (1997?). Traffic and safety, Section 6.1 in: Gartner, N.H., Messer, C.J., and Rathi, A.K. (eds.) Traffic flow theory. U.S. Department of Transportation. Federal Highway Administration. Draft. (To be published as a new edition of Transportation Research Board Special Report 165).

Hedges, L.V. and Olkin, I. (1985). *Statistical methods for meta-analysis.* Academic Press, Orlando, Florida.

Herms, B.F. (1972). Pedestrian crosswalk study: Accidents in painted and unpainted crosswalks. *Highway Research Record 406*, 1-13. Highway Research Board, Washington, D.C.

Hooper, K.G. (1981). ITE Technical Council reacts to the latest RTOR controversy, *ITE Journal*, **31**, 61-62.

Hunter, J.E. and Schmidt, F.L. (1990). *Methods of meta-analysis.* SAGE Publications. London.

ITE (1993). *The traffic safety toolbox: a primer on traffic safety.* Institute of Transportation Engineers, Washington, D.C.

Janusz, K. (1995). Analysis of accident occurrence on the ramps of the Gardiner Expressway and the Don Valley Parkway. M.Eng. thesis. Department of Civil Engineering. University of Toronto. Toronto.

Johannssen, S. (1982). Effektkatalog. (Inventory of road safety measures.) Oppdragsrapport 64. Norges Tekniske Høgskole. Trondheim.

Kelman, L.W. (1977). An analysis of pedestrian accidents and the accident retrieval system in Metropolitan Toronto. M.Eng. thesis. Department of Civil Engineering, University of Toronto, Toronto.

Kendall M.G. and Buckland W.D. (1967). *A Dictionary of Statistical Terms.* Oliver and Boyd, London.

Kleinbaum, D.G., Kupper, L.L., and Morgenstern, H. (1982). *Epidemiologic Research.* Lifetime Learning Publications, Wadsworth Inc., California.

Kulmala, R. (1995). Safety at rural three- and four-arm junctions. VTT Puplications 233, Technical Research Centre of Finland, ESPOO.

Light, R.J. and Pillemer, D.B. (1984). *Summing up: The science of reviewing research.* Harvard University Press, Cambridge, Massachusetts.

Mahalel, D. (1986). A note on accident risk. *Transportation Research Record 1068,* 85-89. Transportation Research Board, Washington, D.C.

Maher, M.J. and Summersgill, I. (1996). A comprehensive methodology for the fitting of predictive accident models. *Accident Analysis & Prevention.* **28**, 3, 281-296.

Maycock, G. and Hall, R.D. (1984). Accidents at four-arm roundabouts. TRRL Laboratory Report 1120. Transport and Road Research Laboratory, Crowthorne.

Maycock, G. and Summersgill, I. (1994). Methods for investigating the relationship between accidents, road user behaviour and road design standards. Annex III in: Ruyters, H.G.J.C.M., Slop, M., Wegman, F.C.M., (eds.). Safety effects of road design standards. SWOV, Institute for Road Safety Research. Report R-94-7.

McGee, H.W. and Blankenship, M.R. (1989). Guidelines for converting stop to yield control at intersections. National Cooperative Highway Research Program Report 320, Transportation Research Board, Washington D.C.

Miaou, S-P. and Lum H. (1993). Modeling vehicle accidents and highway geometric design relationships. *Accident. Analysis. & Prevention.* **25,** 6, 689-709.

Morris, C.N., Pendelton, O.J., Bishop, M.K. and Scaff, C.L. (1989). Empirical Bayes methodology in traffic accident analyses. Paper presented at the American Statistical Association Winter Conference.

Mountain L. and Fawaz, B. (1996). Estimating accidents at junctions using routinely-available input data. *Traffic Engineering and Control,* **37,** No. 11, 624-628.

News (1985). Road edgelining and accidents. *Traffic Enginering and Control.* **26,** No.2, p.85.

Nash, J.C. (1990). *Compact numerical methods for computers: linear algebra and function minimization.* Bristol, New York, Hilger.

Ogden, K.W. (1996). *Safer roads: a guide to road safety engineering.* Aldershot, Avebury.

Ontario (1992). Ontario road safety annual report. Ministry of Transportation, Safety Research Office. Downsview.

Pearson, K. (1904). Report on certain enteric fever inocculations. *British Medical Journal,* **2,** 1184-1203.

Persaud, B. and Hauer, E. (1984). Comparison of two methods for debiasing before-and-after accident studies. *Transportation Research Record 975,* 43-49. Transportation Research Board, Washington, D.C.

Persaud, B.N. (1986). Safety migration, the influence of traffic volumes, and other issues in evaluating safety effectiveness - some findings on conversion of intersections to multiway stop control. *Transportation Research Record 1068.* 108-114. Washington. D.C.

Persaud, B.N., (1987). 'Migration' of accident risk after remedial blackspot treatment. *Traffic Engineering and Control,* **28:** 1, 23-26.

Persaud, B.N. (1992). Roadway safety - a review of the Ontario experience and of relevant work elsewhere. Report PAV-92-02. Ministry of Transport Ontario. Toronto.

Petitti, D.B., (1993). *Meta-analysis, decision analysis, and cost-effectiveness analysis.* Methods for quantitative synthesis in medicine. Monographs in epidemiology and biostatistics, Volume 24. Oxford University Press, New York, Oxford.

Pfundt, K. (1969). Three difficulties in the comparison of accident rates. *Accident Analysis & Prevention.,* **1,** 253-259.

Phillips, G. (1979a). Accuracy of annual traffic flow estimates from short period counts. TRRL Supplementary Report 514. Transport and Road Research Laboratory. Crowthorne.

Phillips, G. (1979b). Accuracy of annual traffic flow estimates from automatic counts. TRRL Supplementary Report 515. Transport and Road Research Laboratory. Crowthorne.

Pickering, D., Hall, R.D. and Grimmer, M. (1986). Accidents at rural T-junctions. TRRL Research Report 65. Transport and Road research Laboratory. Crowthorne.

Quaye K., and Hauer, E. (1993). The use of forecasting models in the evaluation of safety interventions: a theoretical inquiry. In Daganzo (ed.), Transportation and Traffic Theory. pp. 313-332. Elsevier.

Rosenbaum, P.R., (1995). *Observational studies.* Springer Verlag. New York.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Sage, Beverly Hills, California.

Ross, S.M. (1988). A First Course in Probability. Third Edition, MacMillan Publishing Co., New York.

Shinar, D., Treat, J.R. and McDonald, S.T. (1983). The validity of police reported accident data. *Accident. Analysis & Prevention,* **15,** 175-191.

Short, M.S., Woelfl, G.A., and Chang C-J. (1982). Effects of traffic signal installation on accidents. *Accident Analysis & Prevention*, **14,** No. 2, 135-145.

Spirtes, P., Glymour, C., and Scheines, R. (1993). *Causation, Prediction and Search.* Springer-Verlag Inc., New York.

Statens Vägverk (1981). *Effektkatalog*. Borlänge.

Surgeon General of the United States, (1964). Smoking and Health. US. Government Printing Office.

Tanner, J.C. (1958). A problem in the combination of accident frequencies. *Biometrika,* **45,** 331-342.

Tourin, B., and Garrett, J.W. (1960). Safety belt effectiveness in rural California automobile accidents: a comparison of injuries to users and non-users of seat belts. Cornell Aeronautical Laboratories, Automotive Crash Injury Research, Cornell University, Buffalo.

Vingilis, E., Salutin, L. and Chan, G. (1979). R.I.D.E: A driving-while-impaired countermeasure program. One-year evaluation. ISBN 0-88868-036-8, Addiction Research Foundation, oronto.

Wachter, K. W, and Straf, M. L. (eds.) (1992). *The future of meta-analysis.* Russel Sage Foundation, New York.

Welbourne, E. (1989). A note on the effectiveness measures for DRL. (unpublished).

Yates, F. (1965-1966). A fresh look at the basic principles of the design and analysis of experiments. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Vol IV, 777-790.

Zegeer, C.V., Hummer, J., Reinfurt, D., Herf, L., and Hunter, W. (1987). Safety effects of cross-section design for two-lane roads. Final Report. FHWA/RD-87/008. Federal Highway Administration, US. DOT, Washington, D.C..

INDEX

Cook, 15
Cook, T.D., 5
Cooper, 167