Journal of Official Statistics, Vol. 17, No. 3, 2001, pp. 435–445

Book and Software Reviews

Books for review are to be sent to the Book Review Editor Jaki Stanley McCarthy, USDA/NASS, Research Division, Room 305, 3251 Old Lee Highway, Fairfax, VA 22030, U.S.A.

Modeling Longitudinal and Multilevel Data
Isaac Dialsingh 435
The Science of Self-Report: Implications for Research and Practice
Frederick Conrad 436
Sampling of Populations: Methods and Applications
Mark Pont and Paul Smith 440
Statistical Science in the Courtroom
Karen Callaghan 441

Kai Snabel and Jurgen Baumert. *Modeling Longitudinal and Multilevel Data.* Mahvah, NJ: LE Erlbaum Associates, Inc., 2000. ISBN 0-8058-3054-5, 297 pp. 65 USD.

This book is a compilation of papers delivered at the Berlin Summer School Conference on Multilevel and Longitudinal Modeling in 1998. These papers were so comprehensive and intellectually stimulating that it was thought that they must be put in book form.

The book has fourteen chapters that try as much as humanly possible to cover the latest innovations in Multilevel Modeling and Longitudinal Data Analysis. There is no specified format in which these topics are covered but each paper is quite comprehensive. In terms of coverage, the following is a partial list of topics that are covered:

* Modeling Longitudinal and Multilevel Data
* Multilevel Analyses of Grouped and Longitudinal Data
* A Two-Stage Approach to Multilevel Structural Equation Models: Application to Longitudinal Data
* Modeling Multivariate Change
* An Introduction to Latent Growth Models for Developmental Data Analysis
* Modeling True Intraindividual Change in Structural Equation Models: The Case of Poverty and Children's Psychosocial Adjustment
* Modeling Simultaneously Individual and Group Patterns of Ability of Growth or Decline
* Latent Transition Analysis as a Way of Testing Models of Stage-Sequential Change in Longitudinal Data
* Testing Cross-Group and Cross-Time Constraints on parameters Using General Linear Models
* Selectivity and Generalizability in Longitudinal Research: On the Effects of Continuers and Dropouts

© Statistics Sweden

* Multiple Imputations in Multivariate Research
* Longitudinal and Multigroup Modeling with Missing Data

Each paper presented not only covers theory but real-life data are often incorporated using popular statistics programs such as LISREL, EQS, MX, AMOS, HLM and MLn. Graduate students would be the ones who would find this book most useful. The book assumes a lot of statistics knowledge and it is not generally intended for a novice. A good reference section is given so that students wishing to do further research can pursue that avenue with relative ease. In conclusion, I thought that the text was especially good since the development of techniques handling longitudinal and multilevel data has been increasing at an alarming rate. Libraries should be encouraged to purchase a copy of this book for their graduate students.

Isaac Dialsingh
Data Trends and Surveys Limited
Dyette Street
Cunupia
Trinidad and Tobago
Phone: 1-868-665-0176
consult@tstt.net.tt

Arthur A. Stone, Jaylan S. Turkkan, Christine A. Bachrach, Jared B. Jobe, Howard S. Kurtzman, and Virginia S. Cain., *The Science of Self-Report: Implications for Research and Practice*. Mahwah, NJ, 2000. ISBN 0-8058-2991-1, 346 pp, 39.95 USD.

People cannot or do not always accurately report their experiences. For example, when experiences are hard to distinguish or happened long ago, research participants are relatively likely to misreport them – even though trying to report accurately. And when activities are potentially sensitive, such as drug use or abortion, participants may be more likely to intentionally misreport. Yet, despite the drawbacks, such self-reports are the lifeblood of survey data. In many cases there is no alternative source of information about respondents and certainly none that is as affordable. For at least a decade, survey researchers have been consumed by efforts to understand and reduce reporting error. Witness the recent stream of publications about measurement error (e.g., Biemer et al., 1991; Lyberg, et al., 1997) and cognitive aspects of survey measurement (e.g. Sirken et al., 1999; Sudman et al., 1996; Sudman and Schwarz, 1996; Tanur, 1992; Tourangeau et al., 2000).

Survey researchers can take some comfort in the fact that they are not alone. Problems with self-report data limit the inferences that researchers can legitimately draw in numerous disciplines, e.g., psychotherapy, forensic investigation, and clinical assessment of bodily symptoms, to name but three. The insight that there is commonality across self-report methods and that there has been substantial interdisciplinary progress in reducing reporting error served as the impetus for a 1996 conference at the National Institutes of Health. The organizers assembled an elite group of researchers to present overviews of the areas in which they specialize. A collection of papers presented at that conference appears as *The Science of Self-Report*.

All of the 19 papers in the volume are thoughtful, scholarly and accessible. More than half of the papers concern reporting error in surveys. Of these, several discuss the psychological origins of reporting error (e.g., difficulty recalling and dating events); several others describe new data collection technologies that increase the reports of (presumably underreported) sensitive behaviors by increasing the privacy of reporting. Several alternatives to traditional questionnaires are also discussed. For example, rather than asking people how often they engage in habitual behaviors like smoking, an effective method is to automatically prompt participants throughout the day and require them to record what they are thinking and doing when prompted.

Several other chapters concern the accuracy of self-reports in methods other than surveys. These include participants' reports of their adherence to drug regimens and patients' reports of pain and other somatic sensations. Two papers review the effects of mood and imagination. Two papers discuss general issues such as their advantages and ethics.

The most likely benefit to readers involved in producing statistics will be a deeper appreciation of the limits of self-reports as data. For example, if people's ability to report an experience as visceral as bodily pain is affected by how fulfilled they are by their job then surely survey respondents' ability to report less distinctive, less memorable experiences is at least as subject to error. Yet readers might be unable to immediately apply the techniques and insights from domains outside survey research to the reduction of survey measurement error.

Baldwin's chapter begins the book by pointing out that the validity of self-reports is inherently limited. She argues that researchers who use such data should be explicit about how much error they can tolerate and how much precision they require. The desired levels of precision may be impossible if the requisite information is just not encoded by participants.

Tourangeau makes crystal-clear that the degree to which people encode events directly limits the quality of their self-reports. In one study, parents were barely above chance levels of accuracy answering about their children's vaccinations moments after the vaccinations were administered because, presumably, the events were poorly encoded. He also points out how retrieving information about an event, once encoded, may be compromised by changes in storage over time and interference from similar events.

One kind of information that people cannot accurately retrieve is the dates of events. Bradburn attributes much of dating error to the way events are organized in autobiographical memory – narratives rather than veridical, time stamped records. Bradburn discusses eight findings about dating accuracy that practitioners will surely find useful.

Closely related to determining when events happened is determining how often they have happened – their frequency. Menon and Yorkston explore some of the factors that affect the accuracy of frequency reports. They specify when behavioral frequency responses are likely to be based on counts of recalled events, rates of occurrence, and information inferred from a response scale.

Kihlstrom, Eich, Sandbrand, and Tobias discuss the effect of mood on recall. Mood matters when people experience events (e.g., happy events are more accurately recalled than sad ones) and at the time of recall (e.g., people generally recall more accurately when they are happy than sad). The degree to which these match also affects the quality of recall (e.g., people are better at recalling sad events when they are sad).

Schaeffer explores the influence of memory error and impression management on self-reports of sensitive behaviors. She reviews several studies that demonstrate reporting error in socially desirable directions for topics like number of sexual partners, marijuana use and abortion, though in some cases the error is reduced when respondents listen to questions over headphones and directly enter their responses (Audio Computer Assisted Self-Interviews or ACASI).

Reporting about sensitive topics may vary from group to group. Miller, Gribble, Mazade, Rogers, and Turner illustrate how this can lead to unwarranted inferences about other differences between groups. In particular they argue that a widely cited association between abortion and breast cancer may have little to do with causality but, rather, more accurate reporting of abortion by women who have breast cancer than by women who do not.

Respondents who engage in illegal activities may well be reluctant to report their behavior, and, even if located, may refuse to be interviewed. Massey studies such populations by supplementing traditional surveys with ethnographic methods, incorporating the benefits of sampling with the richness of qualitative research.

Another challenging population to survey is children. Blair discusses a pretesting project that collected reports about dietary intake directly from children, rather than from teachers or parents, typically considered more reliable. In some cases the children's reports were more complete than their parents', suggesting they can provide valid self-reports and useful information for pretesting survey instruments.

Drug dependent populations are also hard to survey. Anthony, Neumark, and Van Etten reach such respondents using: (1) ACASI, (2) an item sampling technique to calculate probability of drug taking behaviors without requiring explicit admission, (3) ethnographic techniques, and (4) experience sampling in which participants are beeped at random intervals and required to record their thoughts and actions at that moment.

Shiffman has used a computerized experience sampling method (the electronic diary) to study smoking addiction which he argues is more veridical than the retrospective reports provided as answers to conventional survey questions. The latter are often inaccurate because addicted respondents' answers on the basis of prototypical patterns.

Another mechanical method of capturing behavior is automatic monitoring of medicine use. Rand compares the data from this method to self-reports of adherence to medicine taking regimes and concludes that people chronically overreport adherence. This can compromise evaluation of a drug's effectiveness and the development of dosing guidelines.

Overreporting can occur when imagined events are confused with events that are actually experienced. Loftus describes several experimental demonstrations of this phenomenon: participants are asked to imagine engaging in events and this seems to increase erroneous recollection of the events. She proposes that, in surveys, respondents may answer by imagining themselves carrying out the activity described in the question.

Loftus' work has been interpreted as evidence that some reported sexual abuse is actually imagined even though she discusses underreporting of real abuse. Williams, Siegel, and Pomeroy document such underreporting. They demonstrate that many women whose records show they had been abused would be labeled non-victims if only asked whether they had been abused. Additional questions about the abusive event led a large number of victims to eventually report the abuse.

Kessler, Wittchen, Abelson, and Zhao present a collection of methods that seem to increase the valid report of psychiatric disorders by survey respondents. These include decomposing questions, providing specific instructions on how to carry out the response task, and increasing respondents' motivation and commitment.

Pennebaker presents intriguing evidence that even reports of physical symptoms can depend on non-somatic conditions. For example, people who have boring jobs or live in rural settings report more symptoms than their counterparts. Pennebaker provides a checklist to help physicians interpret self-reports of physical symptoms.

Along similar lines, Keefe presents evidence that certain self-reports of pain may be more frequent and reliable in the presence of certain psychological states. For example, patients anxious about a painful dental procedure rate remembered pain better than did less anxious patients. People report more pain when such reports are socially acceptable and less pain when such reports are discouraged.

Patients' physical states may also affect their reports of physical symptoms. Barsky describes research in which palpitation patients are more accurate in detecting palpitation and resting heartbeat than non-patient volunteers. While this heightened sensitivity seems to be purely physical in origin, reports of discomfort by patients also had clear psychosocial antecedents.

While self-reports do indeed seem fraught with error, Bersoff and Bersoff argue that their use may involve ethical tradeoffs without easy answers. They point out, for example, that survey researchers may uncover evidence-warranting intervention (e.g., a respondent reports suicidal tendencies) but rarely have intervention plans. Proxy responses are often provided without consent from the party whose behavior is being reported. Yet fully apprising participants of their potential costs may lower participation rates.

Overall, *The Science of Self-Report* details the error and, thus, the costs inherent in using self-report data. Despite focusing on error, the volume leaves the reader with a sense that through careful, empirical investigation, the properties of self-reports can be better understood and the data used more judiciously.

References

Tanur, J.M. (ed.) (1992). Questions about Questions: Inquiries into the Cognitive Bases of Surveys. New York: Sage, 154–169.

Beimer, P., Groves, R., Lyberg, L., Mathiowetz, N., and Sudman, S. (eds.) (1991). Measurement Errors in Surveys. New York: John Wiley and Sons, Inc., 1–25.

Sudman, S., Bradburn, N., and Schwarz, N. (1996). Thinking about Answers: The Application of Cognitive Processes to Survey Methodology. San Francisco: Jossey-Bass.

Tourangeau, R., Rips, L., and Rasinski, K. (2000). The Psychology of Survey Response. New York: Cambridge University Press.

Frederick Conrad
Office of Survey Methods Research
U.S. Bureau of Labor Statistics
Washington, DC, U.S.A.
Phone: +01 202-691-7513
Fax:+01 202-691-7426
Email: conrad_f@bls.gov

Paul S. Levy and Stanley Lemeshow: *Sampling of Populations: Methods and Applications*. Third edition. Wiley, New York, 1999. ISBN 0-471-15575-6, xxxi + 525 pp. 89.95 USD.

Sampling techniques are widely used in many organisations wishing to collect and publish statistical information. Notably, governments, marketers and the medical world rely on information collected and processed from samples. The third edition of Levy and Lemeshow's book provides a broad and accessible introduction to the standard techniques for carrying out sample surveys.

The book is split into three distinct sections. In the first, some basic concepts are introduced, including the use of sample surveys, population and units, sampling distributions and bias, variance etc. Part two of the book covers the major sampling methods and related issues – simple and stratified random sampling, cluster and two-stage sampling, ratio estimation (including a cursory two-page summary of regression estimation) and methods for variance estimation. Part three of the book then describes a miscellany of various topics related to sample surveys: imputation for nonresponse and telephone sampling in some detail and many others, for example synthetic estimation and design-based estimation, more briefly.

The book's introduction is comprehensive, outlining a range of basic concepts and terms (including examples of parameters that one might wish to estimate from sample surveys) and the reasons for conducting sample surveys. Important equations in the introduction, and indeed throughout the whole text, are conveniently summarised in boxes in the relevant sections. In general the notation and the language are standard, although both have some quirks that might confuse readers who are used to other texts. Examples include the use of *reliability* instead of *precision*, and V instead of cv for coefficient of variation.

A key feature of the book is the very readable style in which it is written. Each chapter begins with a brief introduction, and ends with a summary of the main points. In between, the development of the ideas is logical and proceeds at a steady pace, making the subject matter easy to follow. The text contains myriad examples, primarily from the medical world. Although examples are abundant, the range of examples, while clearly dear to the authors' hearts, is rather narrow. In some cases this can lead to the omission of some major sampling problems. For example there is a good explanation of why simple random sampling is rarely used, but it misses the problem which stratification is most used to solve – having survey variables with highly skewed distributions. In keeping with the style of the written text, there is a tendency to demonstrate results by numerical example rather than by proof, but the reader is referred to more advanced texts if this more technical information is wanted.

In addition each chapter has a wide range of example problems. A useful feature of the book is the inclusion of a range of illustrative examples using both *STATA* and *SUDAAN*. Sample datasets can be downloaded from the publisher's Internet site so that the examples can be used by readers to help build the connection between the methods and the software. A highly commendable feature of this book is that each chapter ends with a comprehensive bibliography.

There is a range of subjects connected with sampling that has been excluded. As with many other books on sampling, there is no chapter or section on sampling frames – Levy

and Lemeshow confine this to a single paragraph. This does not do justice to the fundamental importance to survey statistics of having a good-quality sampling frame, nor the complexities involved in setting up and maintaining a sampling frame. Part three of the book includes brief descriptions of many of the more involved methods in sampling including small area estimation and imputation, but these are only examples of methods from a wider range, and the reader who really needs to know about these topics will need to look elsewhere too. As a text designed to raise these issues for people who may not have a large experience of sample surveys, it would have benefited by the inclusion of similar sections on: questionnaire design and data collection methods; data validation and editing; and, identification and treatment of outliers.

In summary, the book is an extremely readable introduction to the major concepts of statistical sampling. It is likely to find a suitable audience with those who have some statistical training and need occasionally to conduct sample surveys, or for someone learning about sampling with an application in mind (especially if it is medical!). The text focuses on the main aspects of sampling theory, discusses a few (but by no means all) of the practical difficulties of conducting sample surveys, and leaves detailed derivations and theoretical comparison to other texts. However, it covers all the major issues in an accessible fashion, and will help to explain the mysteries of sampling to a wide audience.

Mark Pont and Paul Smith
Business Statistics Methods Team
Office for National Statistics
NEWPORT
NP10 8XG, UK
Phone: +44 1633 812189
Email: mark.pont@ons.gov.uk

Joseph L. Gastwirth (ed.): *Statistical Science in the Courtroom.* Springer-Verlag, New York, 2000. ISBN 0-387-98997-8, 443 pp. 59.95 USD.

R. Bellman (1961), a scientist from the Rand Corporation, noted ''the fault in so many statistical analyses is not so much in sinning as in the lack of realization that one is sinning, or even the lack of acknowledgment of any conceivable type of sin'' (p. 15). Bellman's warning surely applies to the legal setting. In criminal and civil court cases, the importance of accurate statistical evidence cannot be overstated. Yet, many kinds of flaws in the collection and reporting of data combine to create a less than perfect statistical environment. Thus, there is a need for books that consider the role of scientific evidence in the courtroom.

Joseph Gastwirth, renowned professor of statistics and economics and forensic scholar, has collected an interesting series of articles that consider how involvement in the area of legal statistics affects statisticians and their research programs. The emphasis in this volume is on the proper interpretation and utility of statistical evidence, especially that of a probabilistic nature, and how the legal context shapes the questions that statistical experts address. Twenty-two chapters by leading experts profile a variety of court cases

including illicit drug trafficking, employment discrimination, minority representation, the death penalty, product liability, and other litigation. Many of the contributors to this volume have participated in the cases under study, serving as statistical experts or consultants, advising courts, lawyers, and defendants; thus, their multiple perspectives are invaluable. All contributors convincingly show how the application of statistical science in the courtroom can be improved.

In the first chapter, Colin Aitken explores the role of probabilistic ideas in the legal process. At issue is the accuracy of ''match'' probabilities: the notion that two individuals selected at random from some population match on some characteristic (e.g., blood phenotype). Aitken proposes a coherent procedure for considering the ''value of evidence'' based on the likelihood ratio approach. However, Aitken administered a survey to forensic scientists and lawyers and found that few of them clearly understood this technique. In the final section, Aitken uses Bayesian estimation techniques to determine the sample size needed to reliably estimate the proportion of a large consignment of materials that contain an illegal substance.

Alan Izenman addresses relevant issues in applying federal sentencing guidelines, including sampling procedures and the problem of selection bias. Izenman explores four cases involving illicit drug trafficking, internet pornography, Medicaid fraud, and copyright infringement. Since true random samples of the quantitative amount in question are not always available, non-probability samples are used. As Izenman shows, however, crucial issues of sampling error and uncertainty are often ignored.

David Balding shows why classical hypothesis testing (i.e., the defendant is/is not guilty), familiar to most scientists, is not appropriate for interpreting DNA evidence. Instead, he uses Bayesian estimation techniques to address the relevant question: Is the accused actually the culprit given the evidence? Balding concludes by discussing probability-based reasoning in the U.K. court system and the problem of tutoring jurors to perform their own Bayesian calculations.

Seymour Geisser considers the threats to accurate scientific reporting in the legal setting. He shows how calculations of the relative frequencies of the ''match profile'' depend on assumptions about statistical independence and the random sampling of individuals from homogeneous populations. He also shows how different ''matching criteria'' used by the FBI and commercial forensic laboratories result in higher false ''inclusion'' rates than previously acknowledged. While no guilty person will evade the DNA net, Geisser notes, some innocent ones will be entangled. Geisser concludes by discussing some of the ''sins'' associated with the peer review process.

Bruce Weir, a prosecution's witness in the O.J. Simpson trial, describes his involvement with the forensic use of DNA evidence. He considers the validity of the independence assumption when there are two victims at the crime scene. Weir also describes how the media misreport DNA evidence by ''transposing the conditional'' – i.e., reporting the probability of guilt or innocence, rather than the estimated probability of finding a match in a randomly selected member of the population. His discussion reveals how easily valid scientific evidence can be distorted by the media.

Three chapters take a different approach to DNA evidence in the courtroom. First, Janet Chaseling describes the difficulties of giving expert testimony in the Australian adversarial legal system where the accused's ''right to silence'' can leave the

prosecution's witness unprepared for the forthcoming statistical evidence. Chaseling explains how to present statistical testimony in a comprehensible form. For example, betting on a horserace like the famous Melbourne cup can serve as a framework for explaining the meaning of probability theory in the courtroom.

Second, Beverly Mellen discusses criminal DNA testing in the context of evidential ''relevance'' and likelihood ratios. She explains why the estimated probabilities of observing ''weak'' or ''strong, misleading'' DNA evidence are useful measures of evidence reliability. Mellen discusses several recent U.S. Supreme Court cases that focus on evidential reliability.

Finally, Andres Stockmarr applies the likelihood ratio to ''nonstandard'' cases such as when two different samples are found at the same crime scene along with two suspects. Since each suspect's case must be treated separately, difficulties arise in both calculating and presenting statistical results. Stockmarr also compares the role of DNA evidence in paternity cases involving immigrant children or legal refugees with DNA profiling in criminal cases.

Marc Rosenblum provides an historical review of the evolution of the statistical issues in employment discrimination cases over the last 25 years. He focuses on the changing nature of analytic proof and expert opinion and the factors responsible for this change. Rosenblum stresses the gate keeping role of trial judges who scrutinize expert reports so that juries hear only scientific evidence that is suitable for their fact-finding mission.

David Pollard analyzes the jury selection process in the Connecticut superior court system. According to the state's expert, Hispanics often fail to appear among qualified jurors due to their ''non-participatory'' behavior rather than their systematic exclusion from the jury-selection process. That is, the failure to appear correlates with greater mobility, non-citizenship, and a lower inclination to vote or take out driver's licenses. However, Pollard separates out the question of what happened to potential Hispanic jurors on the ''source'' list from what happened to them after the ''master'' list was constructed and finds a higher rate of undeliverable summonses among Hispanics compared to whites. Outlier methods reveal the unusually large influence of the city of Hartford in the state's regression model, which he sees as poorly constructed.

Joseph Gastwirth considers the statistical issues involved in racial discrimination cases, especially the problems statistical experts face in obtaining relevant data. He stresses the importance of involving statistical experts in the discovery process, and proposes a series of procedural rules that would allow parties to exchange the data and statistical analyses they intend to use at the trial. Gastwirth also reviews the pupil-teacher ratio measure he developed for determining whether school systems are segregated.

Charles Mann provides a practical glimpse into statistical consulting in the courtroom. He contrasts the scientific world where researchers seek ''truth'' about natural law with the courtroom environment where, based on the ''advocacy principle,'' clients' interests are presented in the best possible light. Mann explains the problems that emerge when the rules of constraint on the reporting of scientific evidence are pre-established and administered by legal, non-statistical, experts.

Four chapters in this volume focus on specific policy cases. First, Sana Loue considers the use of epidemiological evidence in toxic torts litigation where injury is alleged to have resulted from exposure to a specific substance or produce. She discusses the legal

procedures for determining the cause of injury and describes various measures of risk, illustrating how they are applied in different court cases.

Next, Wendy Wagner evaluates the standards of judicial review in the Environmental Protection Agency's statistical analyses of environmental rulemaking. Statistical weaknesses such as the inadequate characterization of variability and uncertainty are discussed in light of the courts' demand for ''certainty.'' The author presents a well-reasoned proposal for reforming the standards of judicial review.

Finally, two chapters consider the role of statistical testimony in the Minnesota v. Philip Morris et al. tobacco case. Experts on opposing sides each take a different approach to statistical modeling. Scott Zeger, Timoth Wyant, Leonard Miller, and Jonathan Samet estimate the smoking-attributable expenditures of the state using a regression model and show how the courts misinterpreted their statistical testimony. Donald Rubin uses a statistical inference approach to estimate the tobacco industry's liability for health-care expenditures. He estimates the difference between Minnesota's actual medical expenditures and those hypothesized to occur in the absence of smoking. By including statistical testimony from both sides of a court case, Gastwirth enhances and refines our knowledge about how statisticians formulate a problem.

In the next chapter, Joseph Kadane traces the development of his theoretical ideas about statistics. He explains how exposure to legal issues framed new ideas for his own research agenda. Kadane discusses the strategies lawyers should consider when selecting juries, the optimal use of peremptory challenges, and how the legal system deals with the issue of fairness. He also reviews the ''multiparty'' Bayesian technique he developed, which considers how groups of individuals might look at the evidence.

Nozer Singpuralla explores the issue of product reliability and the probabilistic ideas that underlie the type of warranty contracts that deal with the exchange of a large collection of items. Such contracts are written in terms of ''observable'' quantities – i.e., the number of failures within a prescribed time. They specifically avoid the word ''probability.'' However, Singpurwalla shows how subjective, objective, and logical or ''equilibrium'' probabilities provide a scientific basis to a warranty contract. In the tradition of economics, he calculates the expected losses to buyers and sellers using this approach.

Robert Cottrol addresses one of the most salient controversies in the modern western world: whether the death penalty deters murder and other serious crimes. He reviews two major quantitative studies that have shaped the legal and policy debate on deterrence. This article is informative for what it reveals about the omission of scientific evidence in the courtroom.

The last three chapters of the book consider the Shonubi drug sentencing case. First, Alan Izenman reviews the technical aspects of the case. The only evidence of drug smuggling was found in the balloons the defendant swallowed on his last trip from Nigeria. Should estimates of what Shonubi carried before this trip be based on what drug runners in similar cases had carried, as the prosecution argued? The two chapters that follow provide an affirmative answer to the question and show how the courts failed to appreciate the statistical evidence presented in the case. Joseph Gastwirth, Boris Friedlan, and Weiwan Miao show how missing data, omitted variables, and the violations of other statistical assumptions (e.g., normality) impact drug sentencing. Similarly, Alan Izenman discusses the issue of specific evidence pertaining directly to the criminal

defendant. He shows how additional information would have produced more accurate confidence intervals for the estimated total amount of heroin. The author places a heavy ''burden of proof'' on critics of statistical evidence. He argues that statisticians who identify flaws in the construction of confidence intervals must be required to show how these flaws significantly change the statistical inference.

One weakness is this book's unevenness in style, difficulty level, and article length. Furthermore, since this is a volume of considerable breadth, a more thematically oriented preface was needed to help the chapter contents flow more smoothly. Another problem is that topics like ''the prosecutor's fallacy'' appear repeatedly. Finally, the reference sections at chapter ends fail to follow a uniform style of presentation.

Despite these criticisms, the book is an excellent resource. It brings to light the most current ethical and statistical debates in the courtroom. Thus this book will be of great interest to forensic scholars, statistical experts and consultants engaged in legal proceedings, judges who preside over trials with scientific evidence, geneticists, epidemiologists, and students of statistics and criminal justice, among others.

References

Bellman, R.E. (1961). Adaptive Control Processes: A Guided Tour. Princeton, NJ: Princeton University Press.

Karen Callaghan
Department of Political Science
University of Massachusetts, Boston
Boston, MA 02125
U.S.A.
Phone: 617 287-6947
Fax: 617 287-6511
Email: Karen.Callaghan@umb.edu