



Taylor & Francis  
Taylor & Francis Group



---

Review

Reviewed Work(s): Incomplete Data in Sample Surveys. by William G. Madow, Harold Nisselson, Ingram Olkin and Donald Rubin

Review by: D. M. Titterington

Source: *Journal of the American Statistical Association*, Sep., 1985, Vol. 80, No. 391 (Sep., 1985), pp. 768-771

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: <https://www.jstor.org/stable/2288499>

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Taylor & Francis, Ltd. and American Statistical Association are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*

## Incomplete Data in Sample Surveys (Vols. 1–3).

William G. Madow, Harold Nisselson, Ingram Olkin, and Donald Rubin (eds.). New York: Academic Press, 1983. Vol. 1: xxiii + 495 pp.; Vol. 2: xxv + 579 pp.; Vol. 3: xxiii + 413 pp. \$50.00 each.

### 1. INTRODUCTION

As the preface to Volume 1 explains, these three books represent the fruits of the labors of the Panel on Incomplete Data, which was set up in 1977 by the Committee on National Statistics within the Commission on Behavioral and Social Sciences and Education of the National Research Council. Their task was "to make a comprehensive review of the literature on survey incompleteness in sample surveys and to explore ways of improving the methods of dealing with it." The panel consisted of nearly 20 eminent experts, under the chairmanship of I. Olkin and with the clear presence (in body in the early stages, in spirit thereafter) of W. G. Cochran, whose birthplace lies within five miles of where this review was written. There is no doubt that according to their terms of reference, the panel achieved its objective, which is hardly surprising given the scale of the exercise and expertise behind it.

Most of the chapters are attributed to specific single and joint authorship, with contributions from over 70 individuals. The panel members themselves provide the corporate 100-page report with which Volume 1 begins. The remainder of Volume 1 describes 10 case studies, Volume 2 provides a full investigation of underlying theory (including extensive bibliographies), and Volume 3 records the proceedings (papers plus discussion) of a symposium held in Washington in August 1979.

The degree of interest in and activity surrounding the problem, as well as its importance, is manifested by the size of the books (nearly 1,500 pages in all) and the extent of the bibliographies in Volume 2. L. V. Hedges and I. Olkin provide a considerable service with their annotated list of nearly 80 key papers, neatly abstracting each into a commentary of up to a page. The contribution of B. Bogeström, M. Larsson, and L. Lyberg is less immediately informative but certainly monumental, taking the form of a reference list of about 1,200 publications. (Future research students in this area will be deeply grateful for release from this aspect of their donkey work!) As we shall see, however, the problem of incomplete data in surveys represents a tantalizing rainbow's end in that for all of the mighty forces that have been brought to bear toward its solution, a wholly satisfactory treatment is essentially impossible.

In the following sections I briefly summarize the various parts of the volumes and finally take a personal view of the state of the art.

### 2. VOLUME 1—THE PANEL'S REPORT

Although the panel's report accounts for 100 pages or so, the payoff comes in the first short chapter. After briefly introducing the problem and summarizing the rest of their report, they present 7 concise conclusions and make 35 recommendations (24 related to survey operations and 11 to future research). What are to become familiar points are made here: nonresponse should be eliminated as far as possible; evaluation of methods for analyzing incomplete samples is rarely possible from the data at hand; and considerable computer-related research should be undertaken.

Chapter 2 introduces the problem, defining undercoverage, unit and item nonresponse, illustrating the variability of nonresponse rates among real surveys, and showing how, typically, the degree of bias eventually incurred depends on the nonresponse rate.

Chapter 3 demonstrates the variety of definitions of nonresponse rates and mentions some difficulties in calculating them.

Chapters 4 and 5 provide very concise but informative résumés of the rest of Volume 1 and of Volume 2. Chapter 4 lists the 10 case studies and describes for each the efforts made to reduce nonresponse and the procedures used for imputation for items that are missing or otherwise unusable (as a result of edit failures, for instance). The chapter concludes with a helpful demonstration of the effect on sampling variances of treating imputed values as if they were valid; see also Section 6 of this review.

Chapter 5 contains a concise but very clear account of the main points of all of the contributions to Volume 2, *Theory and Bibliographies*, and the report concludes with brief biographical notes about the panel members.

### 3. VOLUME 1—THE CASE STUDIES

A helpful feature throughout Volumes 1 and 2 is the existence of introductory overviews and outlines of what is to come. H. Nisselson, for instance, provides a brief but helpful summary of the 10 case studies, with tabular descriptions of the sample designs, data collection units and methods (e.g., mail, telephone, or personal interview), the adjustment procedures for *unit* nonresponse, and the imputation methods for *item* nonresponse. In the separate chapters about the case studies, these features are described in full detail. In many cases extensive results are provided concerning response rates and about the performance of the imputation procedures, in as much as this can be judged. As

is repeatedly stressed throughout the treatise (I wish I had bothered to count how often!), the best way to deal with incomplete data is not to have it (them?), and further, any method of adjusting for the effects of nonresponse is usually based on assumptions that cannot be tested. Thus, when appropriate, mention is made of artifices for increasing response rates (by callbacks and incentives, for instance), but the assessment of the extent to which bias is reduced by imputation is generally based on the assumption that at some level or another, the respondents and nonrespondents are "similar."

The full list of the case studies, identified by chapter number, is

2. ECI: the survey for measuring the Employment Cost Index of the Bureau of Labor Statistics of the United States (B. A. Boyes and M. E. Conlon).

3. LIS: the Livestock Inventory Surveys (hogs and cattle) of the USDA (R. L. Freie, with appendixes by B. L. Ford, who reports comparative studies of various hot-deck and other procedures, and C. P. Gleason with R. R. Bosecker).

4. SCA: the Survey Research Center's Survey of Consumer Attitudes (C. G. Steeh, R. M. Groves, R. Comment, and E. Hansmire). Sometimes over 16 telephone callbacks were used to try to avoid nonresponse.

5. OES: the Office Equipment Survey (R. Valliant, R. Tomassino, and M. H. Hansen).

6. ASM: the Census Bureau's Annual Survey of Manufactures, including a description of the General Edit Program (W. G. Madow). Nonresponse rates are observed to be increasing.

7. SCF: Statistics Canada's Survey of Consumer Finances (R. Turner and M. Lawes). Nonresponse rates are shown to increase with average income.

8. NLS: the National Longitudinal Survey of the High School Class of 1972 (M. H. Rizvi). Hot-deck and weighting class methods for item nonresponse are compared: the former do better with discrete items, the latter with continuous, at least as far as bias reduction is concerned.

9. RTTM: a survey of the Readership of Ten Major Magazines (W. G. Madow).

10. CHAS: the 1970 survey for the Central Health Administration Studies of the University of Chicago (M. J. Banks, R. Andersen, and M. R. Frankel), including a comparison of mean imputation and two hot-deck methods that showed little difference among them.

11. HANES: the Health and Nutrition Examination Survey of the National Center for Health Statistics (D. W. Chapman). A comparison is made of the performance, in terms of bias-reduction, of six weighting class adjustment procedures of varying degrees of complexity. Since there turns out to be little choice among them, Chapman recommends the adoption of the simplest procedure.

### 4. VOLUME 2—THEORY

Mention has already been made of the bibliographies in this volume. The remainder of the volume is split into six parts describing the theory underlying the problem. Most of the parts are prefaced by short, helpful introductions that abstract the constituent papers and generally set the scene.

#### 4.1 Part I: Introduction

Even the introduction has an introduction, this one by A. P. Dempster and D. B. Rubin. A superfluous hierarchy of introductions? I think not. They reiterate the main purposes of the panel's activities and present a useful outline of the volume, explaining the main concepts.

The other contribution to Part I is W. G. Cochran's "Historical Perspective," with reflections on the early work of individuals such as Yates, Hartley, and Deming, and his own notes on the remainder of this volume, with particular attention paid to Part IV.

#### 4.2 Part II: Selected Topics in Data Collection

As hinted earlier, all efforts must be made to avoid having to deal with incomplete data, and this part is devoted to this aspect of design. The introduction by J. Sedransk summarizes the various contributions, which discuss *callbacks*, *follow-ups*, and *repeated telephone calls* (P. S. R. S. Rao), *substitution methods* (D. W. Chapman), *quota sampling* (B. F. King), *randomized response techniques* (L. Emrich), and *network sampling* (M. G. Sirken). Rao describes Deming's model for callbacks, Chapman discusses several studies of the effects of using substitution *in the field* (not to be confused with post-sampling "donors" of imputed values), Emrich points out how randomized response techniques can encourage response and yet allow answers to "sensitive" questions to be assessed, and Sirken explains how network sampling links potential nonrespondents with knowledgeable informants.

#### 4.3 Part III: Nonresponse and Double Sampling

As the introduction by W. G. Madow and B. Singh explains, this part considers the possibility of reducing nonresponse, and therefore nonresponse bias, by taking a second sample from those who failed to respond to the first sample. There are two main, fairly technical chapters. P. S. R. S. Rao uses the randomization approach, with estimation of a finite population total in

mind. Estimators, biases, variances, and optimal sample sizes are discussed. A model-based Bayesian approach is adopted by Singh, for structures roughly similar to those considered by Rao.

#### 4.4 Part IV: Weighting and Imputation Methods

This part kicks off with a much longer introductory chapter than usual. In it, D. B. Rubin establishes concepts, notation, and inference methodology for a univariate quantity. First he compares randomization and Bayesian (model-based) inference without nonresponse, noting that in practice, the two approaches often lead to more or less the same inference. In fact, both approaches require a model for the design  $D$  that represents the sampling indicators. Similarly, if there may be nonresponse, both approaches require a model for the distribution of  $M$ , the set of missing value indicators. If  $Y$  represents the set of population values of the variable of interest, it is natural to provide models for the conditional “density” as  $P(D | Y)$ , as far as the sampling mechanism is concerned, and for  $P(M | D, Y)$  to represent the nonresponse mechanism. In a completely model-based approach, a “superpopulation” form for  $P(Y)$  is also suggested so that, altogether, the joint distribution of  $M$ ,  $D$ , and  $Y$  can be constructed in the form

$$P(M, D, Y) = P(M | D, Y)P(D | Y)P(Y). \quad (1)$$

From this, in principle, inference can be made about the missing values within  $Y$  by suitable conditioning, and about any parameters in the model by using the likelihood

$$\int P(M, D, Y) dY_u, \quad (2)$$

where  $Y_u$  denotes the “unknown” values within  $Y$ , whether unsampled or missing.

The form (1) is important because it identifies the stumbling block to most simple procedures for dealing with incomplete data: the need to specify a reliable model for  $P(M | D, Y)$ . Usually, probability sampling of some sort is carried out so that  $P(D | Y)$  is independent of  $Y$  and is known. On the other hand,  $P(M | D, Y)$  is usually not known, and may well depend on  $Y$  and *in particular* on nonrespondents’ values within  $Y_u$ . How therefore can any model be shown to be reliable? The answer is that unless the nonrespondents’ “true” values can be discovered, it cannot! As Rubin points out, the most sensible, pragmatic procedure is to undertake a sensitivity analysis, in which inferences are drawn under a variety of plausible models for  $P(M | D, Y)$ . If the resulting inference is robust as to the choice of model, then well and good; if not, then further fundamental (and usually inaccessible) information is required before any reliable inferential statement can be made.

We shall refer to this infuriating state of affairs again. It is often skipped over, in that many techniques assume that “the missing data process is ignorable” or “respondents and nonrespondents are similar, to a greater or lesser extent.” I have spent a comparatively long time discussing this chapter by Rubin, but justifiably so, I feel, because it summarizes the most tantalizing feature of the whole problem.

Rubin also briefly summarizes the other chapters in Part IV. In the first, H. L. Oh and F. J. Scheuren describe weighting adjustment procedures for unit nonresponse. In their simplest forms these effectively impute sample means for missing values. Probability sampling weights and complicated adjustment cell structures may be involved, but at some level it is necessary to assume that “respondents and nonrespondents are similar.” A wealth of information is provided about biases, bias-reduction, and mean squared errors of estimates, as well as raking and subsample weighting.

B. L. Ford reviews hot-deck estimators, describes three working examples, and summarizes several empirical, comparative studies. Usually hot-deck procedures reduce bias but increase variance, so it is not surprising that he is unable to pick out an overall “best buy”!

We are back to Rubin for the final chapter, this time in tandem with T. N. Herzog; they give an account of their multiple imputation approach. In a model-based structure it is possible to simulate values for the missing quantities. If  $m$  independent sets of simulated values are generated, then together with the known values  $Y_k$ , they create  $m$  pseudocomplete samples. From these one can study the sensitivity of inference to different models and obtain more reliable confidence statements about features of interest. Full details are given of the calculations and application to the prediction of Social Security Benefits for certain individuals in the 1973 Current Population Survey.

#### 4.5 Part V: Imputation Methodology: Total Survey Error

This is a two-man tour de force, with the introduction and two main chapters all written jointly by R. Platek and G. B. Gray. They consider estimation of a scalar quantity using one of four imputation techniques:

1. weighting method in adjustment cells
2. duplication in adjustment cells
3. historical or external data substitution for missing data
4. zero substitution method

Estimators (in the Horvitz–Thompson form) are proposed, and the components of their biases (response and imputation) are calculated, along with the several components of their variance (sampling variance, simple response variance, correlated response variance, and others). Complete details are provided in the first main chapter (I have to confess that I did not check them all!), and a hypothetical example is worked through in the other. Finally, several appendixes add more details and a useful glossary of terms that should have been granted a more prominent position in the complete treatise.

#### 4.6 Part VI: Superpopulation Models for Nonresponse

Following the Platek and Gray double act, we have the one-man show provided by R. J. A. Little. The basis for Little’s introduction and two main chapters was laid by Rubin at the beginning of Part IV with his model-based structure. If the superpopulation approach is acceptable, then familiar concepts such as likelihood can be invoked, along with a wealth of well-developed maximum likelihood methodology. The fundamental formula is (2). How easily it can be used depends on the answers to two questions.

1. Do we know all of the components of (1),  $P(M | D, Y)$  in particular?
2. If not (the answer is usually “No”), does it matter?

In other words, can we ignore the nonresponse mechanism?

Little points out the parallel between sampling and response–nonresponse phenomena. Often the sampling procedure is ignorable in that  $P(D|Y)$  is independent of  $Y$ , giving

$$\int P(M, D, Y) dY_u = P(D) \int P(M | D, Y)P(Y) dY_u. \quad (3)$$

In superpopulation modeling, interest centers on the estimation of parameters within  $P(Y)$  or on the prediction of values for  $Y_u$ . The latter can be generated from estimates of  $P(Y)$ , so the former activity is crucial. Suppose that  $P(Y) = P(Y | \theta)$ . It is usually acceptable to assume that  $P(M | D, Y)$  is independent of  $\theta$  so that, as far as likelihood (or Bayesian) inference is concerned, the answer to question 2 depends on the answer to the following.

3. Does  $P(M | D, Y)$  depend on  $Y_u$ ? The answers to 2 and 3 are typically the same. If they are both “No,” the right side of (3) is

$$P(D)P(M | D, Y_k) \int P(Y | \theta) dY_u \propto P(Y_k | \theta),$$

the “natural” marginal density for  $Y_k$ , the known data, obtained from  $P(Y | \theta)$ . (For reliable *repeated sampling* inference, “ignorability” further requires that  $P(M | D, Y_k)$  be independent of  $Y_k$ .)

For the ignorable case, we may invoke familiar maximum likelihood procedures for dealing with incomplete data. Little surveys these, describes the EM algorithm, and treats in detail the cases of multivariate continuous and discrete data with missing values. In his final chapter, Little revisits these two examples to illustrate the possibilities and difficulties in the nonignorable case. A model has to be specified for  $P(M | D, Y)$ , and different models lead to different inferences, but the validity, or otherwise, of a model is unverifiable from the data at hand, in spite of the fact that estimation of unknown parameters within the model is often possible.

### 5. VOLUME 3—PROCEEDINGS OF THE SYMPOSIUM

The proceedings recounted in Volume 3 reflect the constitution of the symposium in the form of eight sessions, in each of which refereed formal papers were supplemented with contributions by discussants. The whole volume is prefaced with an introduction by W. G. Madow.

#### 5.1 Session I: Background

M. N. Murthy sets the scene by defining four crucial “populations” (target, frame, survey, and inference) and recounting experiences with incomplete data in the Asian and Pacific region. Norwegian experiences are then described by I. Thomsen and E. Siring. They point out the current decline in response rates, outline factors affecting nonresponse and some methods for reducing its effects, and construct and test a model for response in surveys with callbacks. In his discussion, R. M. Groves stresses the difference between *nonresponse* and *undercoverage*, the latter often being the more important problem.

#### 5.2 Session II: Aspects of Design

This session concentrates on the aim of “prevention” of the ailment of nonresponse and overlaps somewhat with Part II of Volume 2. L. R. Frankel and S. Dutka emphasize the importance of specification of objectives, questionnaire design, and interviewing strategy. Like Thomsen and Siring, they propose a model for nonresponse under a callbacks regime and fit the model successfully to a particular survey. A factor contributing to the phenomenon of nonresponse is the interviewee’s concern about invasion of privacy. To reduce this concern, *informed consent* is often made a requirement of data collection. T. Dalenius discusses legislation and the practice of informed con-



sent in Sweden and the United States. He also suggests that in statistical investigations, the need for "total" informed consent may well be unjustified. R. R. Hocking is concerned with the respondent burden. He suggests that one might be more successful, in terms of response, if only partial information is demanded of many of the participants in the sample. He proposes a measure of respondent burden and shows how meaningful estimates can be calculated from partial information, essentially through likelihood equations for incomplete multivariate normal data.

### 5.3 Session III: Theory I

The two papers in this section discuss the estimation of finite population totals by using covariate information. In both cases the missing data mechanism is assumed to be ignorable. The main approach of W. L. Schaible is *dependent* on a linear regression model, whereas that in the other paper by C.-M. Cassel, C.-E. Sarndal, and J. H. Wretman, although based on a linear model, reflects the use of probability sampling designs in the calculation of, say, the biases of their estimators. In the discussion, V. P. Godambe worries about "logical difficulties" with the model-based approach, and K. R. W. Brewer reflects on the balance that has to be struck between pure "probability-sampling" procedures and pure "model-dependent" approaches.

### 5.4 Session IV: Computer Methods Assuming Likelihoods

This section reminds us of two strongly related ideas in maximum likelihood estimation from incomplete data, the Missing Information Principle and the EM algorithm. Not surprisingly, M. A. Woodbury outlines the former, and applies it to a model for file-matching. The chapter by V. Hasselblad, J. P. Creason, and A. G. Stead constitutes a tutorial in the EM algorithm for various types of incomplete data. In the discussion D. B. Rubin refers to work by DeGroot and Goel on broken samples (in connection with Woodbury's paper). R. J. A. Little reemphasizes the distinction between ignorable and nonignorable response mechanisms, pointing out that the latter usually obtain in sampling contexts.

### 5.5 Session V: Theory II

As in Session III, only more distinctly, the contrast in approach between classical and model-based procedures is drawn in the two papers here. The problem is that of double sampling and the optimal choice of sample sizes, with a view toward estimating a finite population total. J. N. K. Rao and E. Hughes obtain unbiased estimators and choose sampling plans to minimize variances. B. Singh and J. Sedransk set out detailed Bayesian analyses that incorporate measures of loss based on trade-offs between errors of estimation and sampling costs. In their discussion M. G. Sirken and K. R. Eberhardt provide constructive criticism.

### 5.6 Session VI: A Model and Adjustment Procedure

In Part V of Volume 2, Platek and Gray derived components of bias and variance for estimators of population totals. In the first paper in this session, J. T. Lessler undertakes similar detailed calculations for an expanded model for total survey error. R. G. Jones then compares mail and interview surveys carried out in Canberra by the National Capital Development Commission of Australia. Nonresponse patterns are compared and somewhat unsuccessful attempts are made to adjust for nonresponse bias by using weighting and linear extrapolation. The discussants, R. E. Fay and T. N. Herzog, agree with the pessimistic prognostications of Jones concerning the lack of efficient ways of combatting high nonresponse rates.

### 5.7 Session VII: Issues in Imputation and Public Policy Experiments

B. A. Bailer and J. C. Bailer evaluate and compare, by using probability sampling theory, biases in estimating finite population totals when using "sequential" hot-deck imputation and "equal weights" imputation. The latter corresponds largely to ignoring records with missing values. The whole exercise is univariate, although it is assumed that imputation is carried out within adjustment cells. Thus, provided respondents and nonrespondents in a given cell are "similar," the hot-deck method should decrease nonresponse bias, possibly at the expense of an increase in variance. C. N. Morris discusses general problems with nonresponse in public policy experiments: difficulties in longitudinal surveys, the danger of errors if wrong models are assumed, and the differential effects caused by unequal respondent burdens. He points out the implicit (or explicit) Bayesian nature of all missing-data procedures and reiterates their reliance on unverifiable assumptions.

### 5.8 Session VIII: Hot-Deck Methods and Inference

The final session was devoted to two topics that were singled out for discussion during the course of the symposium.

G. T. Sande, while accepting that the imputation of missing data should be regarded as a "last resort activity," defends hot-deck methods for their practicality, particularly in view of the sophistication now available. I. G. Sande makes the timely reminder that imputation *per se* cannot add new information to the data. Hot-deck procedures were defined as those "where an incomplete response is completed by using values from one or more records . . . from the

same survey." (There was considerable controversy during the symposium as to what a hot-deck method was. Some people thought that the term might include mean imputation, for instance.) Special factors were discussed related to matching (to produce suitable donors) and the evaluation of hot-deck methods. Finally, M. H. Rizvi introduced a 13-person discussion of hot-decks in general. Details of specific procedures were criticized and further steps were taken on the trail to a firm definition of the term.

The remainder of the session concentrated on the general topic of inference from sample survey data and as such was not restricted to the problem of nonresponse. W. G. Madow introduced a paper in which K. R. W. Brewer and C.-E. Sarndal set up a categorization of six approaches to inference, covering a range from full-blooded Bayesian modeling to classical probability sampling and beyond. They favor an approach that is model-based but that relies on randomization inference to provide robustness, and that judges the efficiency of a sample by a criterion derived from probability sampling. (This philosophy is evident in the contribution of Cassel et al. in Session III.) The subsequent discussion is extensive and valuable in revealing current opinions about the topic. The taxonomy of approaches of Brewer and Sarndal is criticized and modified, V. P. Godambe's contribution to the subject is clarified, and D. B. Rubin and R. J. A. Little illuminate what they believe to be the correct Bayesian perspective. Among the other discussants, T. M. F. Smith (having been absent from the meeting itself, he is bemused about the apparent absence of missing data from the discussion) finishes his written contribution with the crucial remark that the *analysis of missing values can only be tackled by a model-based approach*. There is a wealth of food for thought in the discussion contributions, the rejoinders by Brewer and Sarndal, and the afterthought by D. Basu.

This volume (and the treatise) closes with a brief word by T. Dalenius. He remarks on the impossibility of specifying reliable or estimable response probabilities and urges that large-scale simulation experiments be undertaken to assess workable procedures.

## 6. SOME PERSONAL REFLECTIONS

So far as these volumes are concerned, there is no doubt that they provide an excellent account of the current state of the art. Where, however, do they leave us as far as the problem is concerned? "In a pretty impossible position" is one answer if, once the data-collection stage has finished, the data are still incomplete. T. M. F. Smith received universal support in his view that to at least some degree, models are required to deal with the phenomenon of nonresponse, and yet reliance on them constitutes "an heroic assumption" (C. N. Morris, Vol. 3, p. 322). Even treating the model for response as ignorable is "heroic," particularly because it is usually wrong to do so. (The probability that, say, a question about "income" is unanswered is usually dependent on the actual value for income: it is often comparatively high if income is high and sometimes also if income is low.) The problem can sometimes be alleviated by conditioning, for example by creating homogeneous adjustment cells within which to impute and within which respondents and nonrespondents might more plausibly be similar. In the nonignorable case, models for the response mechanism can sometimes be estimated but almost never tested. Greenlees et al. (1982) were able to check their logistic model, but only by discovering the true complete data from external sources. Rubin (1977) created a Bayesian structure in which nonrespondents are allowed to be "different" from respondents, but the prior description of just how they are different is as "heroic" as ever. The highly compelling argument of Rubin and Herzog in Volume 2 is that careful assessment should be made of the effects of a variety of plausible models in the hope that the desired inferences are not sensitive to this brand of uncertainty.

There seem to be considerable difficulties in using imputation techniques securely, quite apart from the great imponderable just discussed. It is fair to say that no single technique is adequate for the wide range of aims for which the data might be used, and it may be necessary to use several purpose-built methods on the same sample to achieve all of the goals.

As I. G. Sande remarked, no imputation procedure can produce new information without external help. It is true, however, that a completed set of data, comprising real and imputed values, might well be used in some later analysis as if it were a genuine complete set. This leads, not surprisingly, to severe pitfalls. Suppose, for instance, that a single quantity is of interest and that the sampling and response mechanisms are truly ignorable. Imputation of the respondents' mean for all missing values creates a sample whose overall total can be used to provide a sensible estimate of the population total. The sample variance of the created full sample will, however, clearly underestimate the "complete sample" variance. If, on the other hand, the respondents' data are used to generate "random" imputations for the missing values, the sample variance is a bit more reliable. There is, however, no point in trying to delude ourselves into believing that a sample that has been completed using respondents' data (by a hot-deck method, for instance) can be expected to behave in all respects as a valid, complete sample. It is very important therefore to flag

values that are imputed, no matter by what mechanism the imputations were produced.

Comparative studies of various approaches in a variety of population structures will be very valuable. It would be very nice to be able to ask for theoretical analysis of realistic multivariate situations, but it is clear from these volumes that even univariate versions of the problem (which account for most of the calculations to date) require very intricate, if fairly elementary, manipulations. It seems therefore that much future work will be based on simulation studies and comparative exercises with specific, real surveys. Given the computer power now available, it seems likely that the model-based, bootstrap-like, multiple-imputation methods will gain support, as in recent work by Rubin and Schenker (1985) on the development of interval estimation procedures.

In spite of the impossibility of the problem, practical necessity will maintain the pressure for further research, and the material in these volumes represents an ideal springboard.

D. M. TITTERINGTON  
University of Glasgow

#### REFERENCES

- Greenlees, J. S., Reece, W. S., and Zieschang, K. D. (1982), "Imputation of Missing Values When the Probability of Response Depends on the Variable Being Imputed," *Journal of the American Statistical Association*, 77, 251–261.
- Rubin, D. B. (1977), "Formalizing Subjective Notions About the Effect of Nonrespondents in Sample Surveys," *Journal of the American Statistical Association*, 72, 538–543.
- Rubin, D. B., and Schenker, N. (1985), "Multiple Imputation for Interval Estimation From Simple Random Samples With Ignorable Nonresponse," unpublished manuscript.

#### Statistical Abstract of the United States: 1984 (104th ed.).

U. S. Bureau of the Census. Washington, DC: U.S. Department of Commerce, 1984. xxviii + 1015 pp. (paperback).

In 1,015 pages, the 1984 *Statistical Abstract of the United States* presents 1,542 statistical tables, supplemented by six appendixes and a front section of tables and charts on recent trends, and accessed by a 39-page subject index. The cover bears endorsements from Daniel Moynihan, James Michener, and Ben Bagdikian, who express, for all users of the *Statistical Abstract*, their admiration of its prudence, breadth, and objectivity, and its ability to change with the world it measures. This 104th edition includes 97 new tables and drops 110 tables that were in the prior issue. Appendix VI presents selected 1980 census data and the 1982 House election data for congressional districts.

The *Statistical Abstract* functions as (a) a reference to statistics on an enormous range of topics; (b) a guide to sources for the data that are shown, and more detailed data; and (c) a lexicon, offering guidance to the concepts that govern collection and presentation of the data. The *Statistical Abstract* claims the first and second of these functions and performs them very well. It performs the third function, that of a lexicon, with uneven success.

An important feature of the *Statistical Abstract* is its relationship to other publications in the *Statistical Abstract* family of publications and, beyond those, to the surveys and statistical programs that supply the numbers presented here. This point is apparent in the *Statistical Abstract's* appendixes. Appendix I tells the reader where in the *Statistical Abstract* to look for the latest figure in series that are published in *Historical Statistics of the United States*. Appendix V tells where to look in the *State and Metropolitan Area Data Book* for state and metropolitan area statistics corresponding to national or state data shown in the *Statistical Abstract*.

Sources of data are indicated in the introductory text to each chapter, in footnotes to the tables, and in Appendix IV. Appendix IV lists for each subject the principal institutions and publications from which the data were drawn. Appendix III, entitled "Statistical Methodology and Reliability," presents short descriptions of the principal federal statistical programs that furnish much of the data shown here. It furnishes an excellent compact overview of 41 major data systems, such as the National Health Interview Survey. Thumbnail characterizations note the universe that is covered, survey design, data-collection procedures, imputations for missing data, examples of sampling error, and brief notations on the completeness of enumeration, response rates, and other sources of bias.

The *Statistical Abstract* does not work so well as a lexicon as it does as a reference and guide to sources. The degree of the shortfall may be seen if we examine a topic on which the *Statistical Abstract* distinguishes itself—namely, metropolitan population. Appendix II presents the concept and statistical definition of the metropolitan statistical area (MSA); a listing of each MSA and its component counties, together with their 1980 population figures; statistics describing the effects on MSA population of changing MSA definitions; and statistics on metropolitan growth attributable to changing boundaries and to growth of population within constant boundaries. (The 1982–1983 issue of the *Statistical Abstract* showed a map of metropolitan areas, too.) In words and

in numbers, abstractly and concretely, by description and by enumeration, Appendix II conveys the metropolitan-area concept.

Perhaps it would be asking too much to expect all topics to receive such splendid treatment. In that case, the book has other models to offer, notably those instances in which it offers help with the concepts by the way it presents the data. For example, early in the chapter on income and wealth, we are given a table (No. 737) that in 18 lines shows the relationships among gross national product (GNP), national income, personal income, and related quantities, and another table (No. 735) showing the relationships among the components of GNP. This is a big help in grasping the concepts. Moreover, it enables us to establish the connections among other, more detailed, tables showing the objects of personal-consumption expenditures or the origins of national income. Tables on the objects and sources of health-care expenditures perform a similar function, showing the relations among components and furnishing a means of connecting the detailed tables in the chapter on health and nutrition. Table 7, on components of population change, shows how births, deaths, immigration, and emigration combine to increase population, and many later tables may be interpreted as various ways of disaggregating (by age, state, etc.) the quantities in that table. The point is that, at a number of places in this volume, well-designed integrating tables furnish a conceptual guide to the subject of the chapter and enable the reader to move with confidence among the tables of the chapter.

No table offers an equivalent integration of the concepts governing statistics on households and families. Those concepts are complex enough to require such a table, and the classification of individuals by family status offers a means of constructing it. Lacking an integrating tabulation, we get very interesting data without the setting that could make transparent their connection to other quantities in the chapter. We are, of course, given verbal explanations in the text. But text can present only one item at a time, and texts usually present concepts without all the corresponding quantities. So we need a "map" that relates the parts to one another while attaching numbers to the concepts. These are the functions of the integrating tables, and there should be more of them. The volume can improve its function as a lexicon by bringing all of the book up to the best practice it has already established.

The functioning of the *Statistical Abstract* as a reference merits some study by the editors. It is puzzling to find data on voting and registration, but not on the political affiliation of the population; such data are available from academic surveys going back to the 1940s. It is puzzling to find data that religious bodies report regarding their finances and numbers of communicants, but none of the excellent survey data on the religious affiliation of the population and the characteristics of religious groups. The absence of such data cannot reflect an aversion to nongovernmental data sources, since such sources are already employed in this book. Nor can this absence reflect the *Statistical Abstract's* commitment to data on the social, economic, and political organization of the United States; few features are more central to our political and social structure. In recent years the Reagan administration has proposed to increase the National Science Foundation's social and economic science budget (which it once slashed), primarily on the grounds that the surveys it supports furnish a continuing data base that is indispensable to federal policymaking. The National Elections Study, the General Social Survey, the Income Dynamics Survey, and kindred programs deserve another look from the editors.

ROBERT PARKE  
National Cancer Institute

#### E. T. Jaynes: Papers on Probability, Statistics, and Statistical Physics.

R. D. Rosenkrantz (ed.). Dordrecht, the Netherlands: D. Reidel, 1983. xxiv + 434 pp. \$49.50.

This book contains 13 articles written by E. T. Jaynes over a span of 25 years (1957–1980). They are reprinted in chronological order, and Jaynes has added introductory comments to each of them.

Having the collection of Jaynes's papers and his comments in one volume is of great value for statisticians. It will help to disseminate Jaynes's ideas, which are not yet well known by most statisticians because Jaynes is (primarily) a physicist and his articles have usually appeared in journals not traditionally read by researchers in statistics.

The articles in the book provide a review of the author's necessarian approach (L. J. Savage's label) to the theory of probability. Jaynes, as the other necessarian statisticians (e.g., H. Jeffreys) do, proposes a subjective approach to inference, based on a nonfrequentist concept of probability, but an objective method for assigning prior distributions. The priors are derived from general principles by logical analysis of prior information. This is viewed by the personalists (e.g., the L. J. Savage school) as trying to get something from nothing. In this book, Jaynes very clearly shows with several examples what you give and what you get with the necessarian position.