Book and Software Reviews

Sarah E. Igo. *The Averaged American: Surveys, Citizens, and the Making of a Mass Public*. (2007). Cambridge, MA: Harvard University Press, ISBN 0-674-02321-8, 408 pp, $35.

In *The Averaged American* Sarah Igo explores the development of new forms of social science inquiry in the 20th century and posits that these research efforts helped shape Americans' conceptions of themselves and the nation as a whole. While this is perhaps an unprovocative argument, the book makes for an interesting journey just the same. Igo focuses on three specific research endeavors that made their way into popular culture. First is a set of two primarily ethnographic studies called "Middletown" (1929 and 1937) that set out to describe a "typical" American city (Muncie, Indiana was the designated site). The second is the public opinion poll, arriving in the late 1930s; Igo focuses in particular on the work of George Gallup and Elmo Roper. And third is the Kinsey Reports on sexual behavior, published in 1948 and 1953. What sets these three projects apart from the types of research that came before them is a combination of new techniques (such as probability sampling), new kinds of inquirers (academics, pollsters and marketers), broader and more public dissemination of findings, and a fundamental shift in the research questions posed. Social science throughout the 19th century generally focused on public health, economic progress and social policies – still the bread and butter of many research agencies today. But these 20th century research projects shifted the focus from social reform efforts, which usually involved the study of marginal populations (such as "social deviants") to questions of the everyday, and of less-observable characteristics, such as what "typical" Americans eat for breakfast, what they think about religion in schools and how they behave in private. Most of the book is an exploration of the ways in which these changes and their subsequent products were received by the public and helped shape the public's perception of itself.

Igo does not attempt to evaluate these new methods in terms of data quality or accuracy. The researchers themselves, however, at times reflect on the relative merits of various approaches. Ten years after the Middletown studies the author, Robert Lynd, offered laments about their methodology, stating that if he had the job to do over again he would

have collected life histories and designed better instruments for measuring attitudes and opinions (foreshadowing persistent concerns about measurement error). Lynd's comment dovetails nicely with Kinsey's approach to his sex studies. In the early stages of his project he was skeptical of prior research in part due to its reliance on questionnaire rather than life history data. But this type of technical commentary is sparse; Igo's main goal is to portray the public's surprisingly active and willing participation in these new research methods, and their simultaneous skepticism and acceptance of the new techniques and findings. Indeed, who knew that "average" Americans would allow ethnographers into their homes, churches, schools, clubs, etc. and willingly expose their everyday lives to probing social scientists? Or that people would be so generous with their time as to answer lengthy questionnaires on toothpaste and/or politicians? Not to mention the unmentionable sexual behavior. Igo draws on a broad range of source materials to produce this portrait – scientific journals, media reports, editorials, letters to the editor and, perhaps most interesting, correspondence between subjects and researchers.

Since this type of portrait is the real meat of the book, readers cannot expect a comprehensive history of the development of the field of social science. There are, however, plenty of important markers along the way and, in the case of the Kinsey Reports, a rather well-developed treatment of the methods and substantive findings. In the Middletown chapters, however, Igo perhaps too quickly jumps into *reactions* to the research – of the city residents as well as the media and general public – which are at times hard to grasp without much discussion of the research findings themselves. Nevertheless, there are some interesting methodological connections across the three research efforts. For example, all explicitly excluded minorities, but in different ways and for different reasons. In the Middletown case the researchers' focus was on native whites for technical reasons, according to the researchers: "Since we are attempting a difficult new technique in a highly complicated field, it is desirable to simplify our situation as far as possible. . .[instead of] being forced to handle two major variables, racial change and cultural change." The Gallup and Roper polls made an attempt to be more inclusive, through the use of probability sampling, but given problems of respondent suspicion and interviewer reluctance to work in certain neighborhoods, the researchers did not always find it worth the trouble and expense. There was a more pragmatic, commercial issue at work as well. For pollsters, election results were a litmus test of the soundness of their methods, so their focus was on the voting public. Since most blacks were barred from the polls, their views were of little interest to the surveyors. Kinsey, on the other hand, collected data from an extremely diverse group of subjects but when it came time for analysis he did not feel he had sufficient cases from minority groups to describe them. Thus the omission of minorities – in the data collection and/or analysis stages – was routine practice. However, there is little in the book to suggest that the researchers, or the public (in the case of Middletown), were very concerned about this omission. And though there are various accounts of criticism for the exclusive focus on whites (e.g., one reviewer renamed the Kinsey Report "Sexual Behavior in the Human Male, White U.S.A.") there is not a very developed discussion of the consequences of the omission of minorities.

A recurrent question is whether these social science methods and findings simply reflected society or whether they actually shaped attitudes, beliefs and behavior. While George Gallup "strenuously denied the existence of such influence throughout his career,"

Roper gave some ground on the question. Critics of Kinsey, on the other hand, feared that his results – on the frequency of behaviors such as homosexuality, extra- and premarital sex – would somehow normalize what they perceived to be abnormal sexual behavior. That is, there was a fear that Americans would take their cues from statistics and not from traditional family values. In spite of these questions, fears, and the methods' shortcomings, Igo makes clear that the public accepted, albeit with some skepticism, these new approaches to understanding society. What is not entirely clear is what role, if any, garden-variety social science (such as that conducted by government agencies, foundations and the like) played in the average American's idea of what "America" was. Research on topics such as disease prevalence, economic indicators and population size were and are part of the national "output" of information available for public consumption, but Igo omits these types of inquiries almost entirely from the discussion and it is not clear why or what impact this exclusion has on the portrait she provides.

While there is no shortage of interesting material about the birth and development of modern social science methods, and how they insinuated themselves into American culture, the book's central argument is somewhat elusive. The subtitle: "Surveys, citizens, and the making of a mass public" presumes the reader is familiar with the concept of a "mass public" but fails to define it or offer a developed discussion of how it relates to ways of knowing. Readers, then, are left to conjure their own notions of what a "mass public" is. So instead of an engaging argument, the book can read as a story where the backdrop is sweeping change in America fueled by a huge wave of immigration, the Great Depression, World War II and the Cold War, among other factors. These major events produced a newfound curiosity and market for information about what "America" was that did not exist in the more atomized, less urban culture of the previous century. The new condition of the country coincided with the emergence and development of new research techniques to describe in great (and sometimes flawed) detail what this society was. So Americans now "know" that 50% of marriages end in divorce, they cannot imagine living in a world where these nuggets of knowledge are not commonplace, and some may even feel intellectually naked without consulting some kind of poll or survey data. These seem like reasonable statements but they are merely posited and not demonstrated; without a clear understanding of what a mass society means, it is difficult to evaluate its state before and after the advent of modern social science.

Overall this book makes for an enjoyable and interesting read. Academics in the humanities may find it a useful source book on the evolution and acceptance of social science methods as valid and in some cases superior ways of knowing. And while it is not a technical treatment of survey methods, practitioners will likely appreciate the history of their field and earlier examples of problems and challenges that are still dogging researchers today.

Joanne Pascale
U.S. Census Bureau
Center for Survey Methods Research
Statistical Research Division, FB4-3000
Washington D.C. 20233
U.S.A.
Phone: 301-763-4920
Email: Joanne.pascale@census.gov

Laurel Schwede, Rae Lesser Blumberg, and Anna Y. Chan (eds.). *Complex Ethnic Households in America*. Lanham, MD:Rowman and Littlefield Publishing, 2006. ISBN 0-7425-4637-3, 336 pp, $29.95.

When we use the terms "family" or "household" we usually refer, whether implicitly or explicitly, to a group of people living together, which is composed of a husband and a wife, often with children. Both American pop culture and "common sense" have, for decades, reproduced the ideal of the nuclear family living in the suburbs, with two cars, and a dog. The reality is, however, that ongoing economic and social changes make it impossible to hold this ideal as the norm anymore. Middle-class wives no longer stay at home taking care of their children – working-class women never did – and the massive arrival of new immigrants from all over the world since the 1960s, together with their struggles to settle in this country, has had a considerable impact on family formation and household structure. If we want to understand what the "American family" looks like today, it is necessary to leave myths behind and to consider the changes in the country's racial composition, as well as in gender dynamics, both inside and outside of the household. It is also important to understand economic processes that may shape decisions around household formation and survival.

This book is a considerable step in the right direction. Through an examination of both qualitative and quantitative data, the authors explore aspects of the ethnic composition of the United States and household formation processes. Most importantly, they analyze various cultural definitions of household that may be at odds with official conceptualizations found at the census. Their main goal is to contrast emic[1] and etic[2] understandings of the household that may ultimately lead to improved questionnaire design and interview training, which could have wider policy implications within the census.

The main body of the book consists of six chapters that discuss findings from ethnographic work – including participant observation, but mostly in-depth interviews- with families forming complex households in six ethnic groups in the United States: the Navajo in Arizona; the Iñupiat of Northern Alaska; Korean immigrants living in Queens, New York; Latino migrants living in Central Virginia; African Americans living in urban coastal Virginia; and rural whites from the upstate New York snow belt.

What the authors found was that, far from reproducing the nuclear family model, these groups are organizing their familial living situation around complex households, which include any household configuration that is not solely a nuclear family. Although the findings from their interviews cannot be generalized to the larger population, they are consistent with trends appearing in the census and in other survey studies. Families from the six ethnic groups organized their households following cultural, economic, and emotional criteria. Trends such as female householder families, skip-generation, multi-household families, and multi-family households appeared in the six communities not only as part of their cultural tradition, in the case of the Navajo and the Iñupiat, but also, and most importantly, as responses to local socio-economic changes and as part of adjustment strategies upon arrival in the United States.

---

[1] The insiders' cultural view.
[2] General definitions in the larger society that may seem to be standard and objective.

This book is an important contribution to the fields of ethnic, family, and gender studies for several reasons. First, it recognizes the importance of gender dynamics in the context of household formation and sustenance. Although it is key to recognize the power dynamics that have historically taken place within families and households along gender lines, the authors take the discussion one step further by connecting gender dynamics with larger issues such as lineage and inheritance. Gender relations within the household, and the power relations stemming from them, are also seen in the context of changing economic opportunities and the latter's impact on women's decisions and leverage within their household. In short, the household does not exist in a vacuum, and it is continuously being transformed by forces operating at micro, meso, and macro levels.

Second, the authors make both traditional and more recent alternatives to the mainstream nuclear family visible, and they demonstrate that the nuclear family is not in crisis, but rather, is being challenged by other models. Rather than seeing these alternatives as an outcome of irrational, backward, or undeveloped thinking, the authors show us that they are, more often than not, a consequence of quite rational and calculated thinking, where families are constantly negotiating: Cultural traditions often become a tool of survival, like in the case of the Navajos. However, they also have to be put aside in order to survive in this country, like in the case of the Latino and Korean immigrants.

Third, this book is a beautiful demonstration of an inherent problem in social research, that is, the clash between the researcher's conceptualizations and the subjects' worldviews. It will be humbling – and illuminating – for any social scientist doing fieldwork to read Nancy Tongue's misunderstandings with her Navajo informants or Tai Kang's ongoing revision of his interviewing techniques. This book will teach the reader much more than complex household formation. It is also a colorful and effective lesson on field research techniques and methods.

Having said this, while reading the book I often wondered whether the authors were referring to ethnic instead of "racial" household. While the ethnic distinction becomes quite clear in the first four cases – and somewhat clear in the fifth, African Americans in Virginia – it does not seem to have any relevance within the group of white families that they interviewed. Although we often forget, whites have ethnicity as well, which is different from our race. The authors do not recognize this. They could have chosen to interview a group of white Italian American or Irish American households in the same area, but instead they chose not to make that differentiation. My main question for them would be – and I realize that it is not an easy question: Why do they emphasize ethnic differences within racial groups in the case of racial minorities but not consider the same differences in the case of white folks? If they were to consider these differences, do they think that ethnicity would be relevant in the case of whites in the context of household formation?

Another issue that I would find interesting to explore further is that of "elderly care." Sharon Hewner's findings – discussed by Jennifer Childs in this volume – indicate the importance of care of the elderly in the context of household formation. Traditionally an unpaid form of work done by the family (read women) care of the elderly and the disabled is becoming in some states, such as California, a form of paid work that is often done by family members. The In Home Support Services program in most Californian counties provides in home care for the elderly and the disabled. This work is often done by family members who, instead of having to choose between working outside of the home or taking

care of a relative, are paid by the county to do the latter. Therefore, working outside of the home often means taking care of a relative. If this program extends to other states, it will surely substantially affect the ways in which households are organized – and how the tension between household and paid work is negotiated – by family members from all ethnic and racial groups.

In a nutshell, some of the issues raised by the authors of this book call for further research and reflection on the usage of analytical concepts by researchers. Despite this, I found this a solid piece of work that teaches a lot about household formation in the United States at the beginning of the 21st century. Its main achievement is, undoubtedly, its ability to challenge census categories as objective and universal, and its argument in favor of a revision of such official categories. In addition to questioning official definitions, it also deconstructs cultural configurations and mainstream perceptions of household formation strategies undertaken by minority groups. Therefore, this book provides tools to understand people's choices from our position as social researchers. However, and perhaps most significantly, it also helps us to understand people, their choices, and their struggles, as human beings. Regardless of how different they or we are.

Sandra Ezquerra
University of Oregon
Department of Sociology
1291 University of Oregon
Eugene, OR 97403-1291
U.S.A.
Email: sezquerr@uoregon.edu

Marcello D'Orazio, Marco Di Zio, and Mauro Scanu. *Statistical Matching: Theory and Practice.* 2006. Hoboken, NJ:Wiley, ISBN 0470023538, 253 pp, $99.

Government agencies and survey organizations collect a variety of data, often on different variables. Integrating such datasets enables rich analyses without expensive new collection efforts. Hence, it is useful for data collectors to have sound theoretical frameworks and practical examples for data integration. The goal of the text by D'Orazio, Di Zio, and Scanu – henceforth abbreviated as DDS – is to provide such frameworks and examples.

DDS use the term *statistical matching* to describe a particular version of data integration. Specifically, they focus on datasets that contain disjoint sets of units and some different variables. Other versions of data integration, not the focus of DDS, include (i) horizontally partitioned data, in which the databases have the same variables but different units, (ii) vertically partitioned data, in which the databases have the same units but different variables, and (iii) partially overlapping data, in which the databases have some records and some variables in common, and others not in common. For a review of these data settings, see Karr et al. (2006). DDS do not address record linkage, in which it is known that some units are in both databases. The statistical matching in DDS is not related to matching in the causal effects literature, e.g., propensity score matching.

As DDS point out, their statistical matching problem is essentially a missing data problem. In Chapter 1, and indeed throughout the rest of the book, they present likelihood

theory for missing data and utilize it to characterize the statistical matching problem. The difficulty for inference, and it is a substantial one, is that giant blocks of information are completely missing. For example, Agency A might have data on (**x, y**) and Agency B might have data on (**x, z**), but **y** and **z** are never observed together. Additional information, typically in the form of distributional assumptions or auxiliary variables, is needed for inference on quantities involving functions of (**y, z**) such as correlations or regressions. DDS note that inferences with such integrated data are sensitive to those assumptions.

The remainder of the book is dedicated to making inferences under potential assumptions. DDS devote Chapter 2 to data integration under conditional independence. Under this assumption, in the example above we would assume that **y** is independent of **z** given **x**. DDS present theoretical results for the multivariate normal and multinomial models, which have special properties that make them easy to manipulate under conditional independence assumptions. They also thoroughly investigate approaches based on hot deck imputations in this chapter, as well as in later chapters.

The conditional independence assumption is unlikely to be true in general practice. To avoid this assumption, data collectors require additional information. Chapter 3 illustrates how various forms of external information can be utilized. When auxiliary data exists for the combined records in the files, DDS use the missing data framework to estimate parameters of the joint distribution, for example with an Expectation Maximization (EM) algorithm. Sometimes the auxiliary data exists on population distribution and not the records themselves. DDS describe likelihood-based and nonparametric approaches for estimating parameters in this setting.

Just as conditional independence is unverifiable (unless external information indicates otherwise), the models used to relate the auxiliary variables with the observed ones may be difficult to check. DDS present approaches to estimating uncertainty in inferences in Chapter 4. For example, it is possible to obtain bounds on the unknown elements from the sample covariance matrix by using the conditions implied by positive definiteness. Bounds also can be obtained for parameters of multinomial distributions. This chapter also discusses algorithms that enable the data collectors to incorporate constraints on parameter values.

Chapter 5 discusses data integration for finite population inferences, when the data collectors may want to use survey weights and design-based estimation. DDS present calibration approaches and file concatenation approaches for this setting. Chapter 6 deals with some preprocessing issues needed to undertake statistical matching, like ensuring consistent variable definitions and subpopulation compositions. DDS also discuss heuristics for determining which variables are worth including in the data integration. Chapter 7 describes an application to a genuine dataset. This chapter provides key guidance for practitioners. The book concludes with appendices on missing data methods, loglinear models, statistical distance functions, and a primer on design-based estimation. Finally, there is an appendix with code in the software package R that implements several of the techniques described in the book, including hot deck imputation, likelihood estimation, and uncertainty estimation.

As a final thought, I believe there are interesting connections between the statistical matching problem and the use of synthetic data to protect confidentiality, as introduced by Rubin (1993). In synthetic data for confidentiality, the objective is to produce a simulated dataset that has the global properties of the original data, but all records are simulated.

The synthetic data are drawn from probability models estimated using the original data. As a consequence, to allow data users to account for the uncertainty due to simulation, the agency should generate multiple copies of synthetic datasets, and apply the combining rules of Raghunathan et al. (2003) to obtain valid inferences. Some statistical matching contexts essentially follow the same logic, for example generating micro data from probability models constructed with information from auxiliary data. One would expect the combining rules of Raghunathan et al. (2003) to hold in this statistical matching situation as well.

In the end, I believe that DDS go a long way towards meeting their objective with *Statistical Matching: Theory and Practice.* By couching the matching problem in the missing data framework, they provide a solid theory for inference. Understanding this theory requires being comfortable with statistical inference and probability-based modeling. Familiarity with missing data methods is also helpful. My one wish is that the book had more genuine case studies besides the one in Chapter 7. Nonetheless, I recommend this book to anyone seeking to learn about the theoretical underpinnings and necessary assumptions in the statistical matching problem.

Jerome P. Reiter
Duke University
Institute of Statistics and Decision Sciences
Box 90251
213 Old Chemistry Building
Durham, NC 27708
U.S.A.
Phone: 919-668-5227
Fax: 919-684-8594
Email: jerry@stat.duke.edu

References

Karr, A.F., Fulp, W.J., Vera, F., Young, S.S., Lin, X., and Reiter, J.P. (2006). Secure, Privacy-preserving Analysis of Distributed Databases, Technometrics (forthcoming).

Raghunathan, T.E., Reiter, J.P., and Rubin, D.B. (2003). Multiple Imputation for Statistical Disclosure Limitation. Journal of Official Statistics, 19, 1–16.

Rubin, D.B. (1993). Discussion: Statistical Disclosure Limitation. Journal of Official Statistics, 9, 462–468.