THE NATIONAL ACADEMIES PRESS

SHARE

The Future of Federal Household Surveys: Summary of a Workshop

BUY THIS BOOK

FIND RELATED TITLES

AUTHORS

Krisztina Marton and Jennifer C. Karberg, Rapporteurs; Committee on National Statistics; Division of Behavioral and Social Sciences and Education; National Research Council

THE FUTURE
OF FEDERAL
HOUSEHOLD SURVEYS

SUMMARY OF A WORKSHOP

Krisztina Marton and Jennifer C. Karberg, *Rapporteurs*

Committee on National Statistics

Division of Behavioral and Social Sciences and Education

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, D.C. 20001

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

STEERING COMMITTEE FOR THE WORKSHOP ON THE FUTURE OF FEDERAL HOUSEHOLD SURVEYS

HAL S. STERN (*Chair*), Donald Bren School of Information and Computer Sciences, University of California, Irvine

KATHARINE G. ABRAHAM, Joint Program in Survey Methodology, University of Maryland

CHESTER BOWIE, National Opinion Research Center, Bethesda, Maryland

CYNTHIA CLARK, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, DC

GRAHAM KALTON, Westat, Rockville, Maryland

JENNIFER MADANS, National Center for Health Statistics, U.S. Department of Health and Human Services, Hyattsville, Maryland

ALAN ZASLAVSKY, Department of Health Care Policy, Harvard University Medical School

KRISZTINA MARTON, *Study Director*
JENNIFER C. KARBERG, *Staff Officer*
AGNES GASKIN, *Administrative Assistant*

v

COMMITTEE ON NATIONAL STATISTICS
2010-2011

LAWRENCE D. BROWN (*Chair*), Department of Statistics, The Wharton School, University of Pennsylvania

JOHN M. ABOWD, School of Industrial and Labor Relations, Cornell University

ALICIA CARRIQUIRY, Department of Statistics, Iowa State University

WILLIAM DuMOUCHEL, Oracle Health Sciences, Waltham, Massachusetts

V. JOSEPH HOTZ, Department of Economics, Duke University

MICHAEL HOUT, Survey Research Center, University of California, Berkeley

KAREN KAFADAR, Department of Statistics, Indiana University

SALLIE KELLER, Science and Technology Policy Institute, Washington, DC

LISA LYNCH, Heller School for Social Policy and Management, Brandeis University

SALLY C. MORTON, Department of Biostatistics, University of Pittsburgh

JOSEPH NEWHOUSE, Division of Health Policy Research and Education, Harvard University

SAMUEL H. PRESTON, Population Studies Center, University of Pennsylvania

HAL S. STERN, Donald Bren School of Computer and Information Sciences, University of California, Irvine

ROGER TOURANGEAU, Joint Program in Survey Methodology, University of Maryland, and Survey Research Center, University of Michigan

ALAN ZASLAVSKY, Department of Health Care Policy, Harvard University Medical School

CONSTANCE F. CITRO, *Director*

Acknowledgments

This report summarizes the proceedings of the Workshop on the Future of Federal Household Surveys, held on November 4-5, 2010. The workshop was convened by the Committee on National Statistics (CNSTAT) of the National Research Council's (NRC) Division of Behavioral and Social Sciences and Education (DBASSE) to discuss major challenges facing the federal statistical system in the area of household data collections and to identify strategies for moving forward.

Support for the workshop was provided by several federal statistical agencies through a core grant to CNSTAT from the National Science Foundation's (NSF) Methodology, Measurement, and Statistics Program. Contributing agencies included the Bureau of Justice Statistics, the Bureau of Labor Statistics, the Bureau of Transportation Statistics, the National Center for Education Statistics, the National Center for Health Statistics, the National Center for Science and Engineering Statistics, the U.S. Social Security Administration, and the U.S. Census Bureau.

As chair of the workshop steering committee, I acknowledge with appreciation everyone who participated in the workshop and made it a success. I especially would like to thank my colleagues on the steering committee for their dedication and leadership in planning the workshop and moderating the sessions. On more than one occasion a steering committee member volunteered to offer their expertise to fill a place in the program. I also thank all of the presenters for their thoughtful presentations and professionalism, and acknowledge the many workshop participants for their contributions. The discussions were bold, and many new ideas emerged that can benefit the federal statistical system.

Hal S. Stern, *Chair*
Steering Committee for the Workshop on
the Future of Federal Household Surveys

Contents

ix

1

Introduction

The Workshop on the Future of Federal Household Surveys was designed to address the increasing concern among many members of the federal statistical system that federal household data collections in their current form are unsustainable. The workshop was held at the request of the U.S. Census Bureau. Other statistical agencies that helped sponsor the workshop through the core grant to the Committee on National Statistics from the National Science Foundation's Methodology, Measurement, and Statistics Program include the Bureau of Justice Statistics, the Bureau of Labor Statistics, the Bureau of Transportation Statistics, the National Center for Education Statistics, the National Center for Health Statistics, the National Center for Science and Engineering Statistics, and the Social Security Administration.

Although no consensus recommendations were generated, the workshop was structured to bring together leaders in the statistical community and to facilitate a discussion of opportunities for enhancing the relevance, quality, and cost-effectiveness of household surveys sponsored by the federal statistical system.

Federal household surveys today face significant challenges: (1) increasing costs of data collection, (2) declining response rates, (3) perceptions of increasing response burden, (4) inadequate timeliness of estimates, (5) discrepant estimates of key indicators, (6) inefficient and considerable duplication of some survey content, and (7) instances of gaps in needed research and analysis (e.g., lack of information on institutional populations). The more recent American Community Survey (ACS) can possibly be leveraged to help cope with these

challenges, and the workshop considered options for better integrating it into the federal household survey portfolio.

Although moving forward with a coordinated strategy will require many more conversations on the issues covered at the workshop, if solutions are not developed in a comprehensive and timely manner, the challenges put at risk the ability of the federal statistical system to provide important policy-relevant information. The goal of the workshop was to begin and to facilitate the much-needed discussion on solutions that range from methodological approaches, such as the use of administrative data, to emphasis on interagency cooperative efforts.

WORKSHOP FOCUS

The goal of and charge to the steering committee were to develop a workshop aimed at enhancing the household surveys sponsored by the federal statistical system. As part of his welcoming remarks, the steering committee chair, Hal Stern (University of California, Irvine) noted three guidelines for participants to keep in mind. First, the workshop was to provide a picture of the system as it is, including an overview of the many current challenges. And although such issues as nonresponse and increasing cost are of great interest, these challenges would be used to set the context for discussion rather than being the focus of discussion themselves, he said, because a number of other recent meetings have focused on these topics extensively.

Second, an important cross-cutting issue was how a large continuous survey, such as the ACS, could be useful to the household survey system. The questions were what could be done with a survey like this and how could it best be used. This issue came with a caution, however, not to get bogged down in the details at this stage of the conversation.

The final caution made by Stern was to avoid the trap of focusing on what cannot be done, which would be the wrong kind of discussion for this workshop. Instead, he emphasized that workshop participants should keep open minds and consider where innovation and experimentation might happen. He said that this was not just a presentation workshop; it was meant to inspire and encourage participation from those present.

Stern said that this point was worth reinforcing: this workshop was intended to be about ideas. It is ever more critical that the statistical community consider ways to make the household survey system better and more efficient. In that spirit, he encouraged the participants to consider some challenging questions. Is the model of data collections centered around individual surveys outdated? How can new data collection modes and analysis techniques be integrated most efficiently? Can the resources invested in maintaining and updating address files be streamlined and perhaps directed toward developing a universal address file?

WORKSHOP ORGANIZATION

The workshop began with a look at the U.S. household survey system and where it stands, followed by overviews of household survey systems from several other countries: the United Kingdom, the Netherlands, and Canada. These countries are facing many of the same issues as the United States. Although what works in one country may not work in another, it is important not to rule any ideas out in the course of these discussions.

The workshop then focused on topic areas in which promising research is being done and there is also room for additional discussion and perhaps some experimentation. One of these topics is sampling frames: Can large surveys serve as first-phase samples for smaller surveys? Can the statistical community work together to make the development and maintenance of sampling frames more efficient? There was also a general discussion of methodology—for example, modes of data collection and the use of administrative records.

The agenda then shifted to a discussion of estimation challenges and the boundaries between direct estimation and model-based small-area estimation. This was followed by a discussion of survey content, particularly instances of multiple measures of the same concept, when this is desirable, when it is not, and what can be done about it. This session included thoughts on the potential future role of the ACS and of the U.S. Office of Management and Budget.

PLAN OF THE REPORT

This summary of a workshop is intended to describe the presentations of the workshop and the discussions that followed each session topic, as outlined in the agenda (see appendix). Following this introduction, Chapter 2 represents the first session of the workshop with an overview of the U.S. federal household survey system at a crossroads. It also presents models of household surveys in other countries in contrast to those in the United States. Chapter 3 covers the session on sampling frames and new ideas on how to use them. Chapter 4 addresses various methods of collection of household data, including the use of administrative records. Chapter 5 summarizes the discussions that took place at the end of the first day's presentations. Chapter 6 covers the topic of small-area estimation, how this methodology is used now, and other ways that it might be used in federal surveys. Chapter 7 focuses on survey content, discussing standardized measures of the same concept used across different surveys (e.g., disability) and instances when the use of different measures is more appropriate (e.g., poverty or income). The chapter also addresses the topic of official statistics. Finally, Chapter 8 summarizes the floor discussion that took place at the workshop's close.

It is important to note that the nature of this report is that of a factual summary of the presentations and related discussions that transpired during the workshop. Therefore, all views presented herein are those solely of the

workshop participants. The presentation topics and content reflect the areas of expertise of the presenters and are not intended to be an exhaustive discourse on the future of federal household surveys. Furthermore, this workshop was not designed to produce either conclusions or policy recommendations. Rather, the intent of the workshop was to open a dialogue on the subject, encourage further research, and share new ideas about improving the system of household surveys.

2

The Federal Household Survey
System at a Crossroads

To set the stage for the workshop, the first session provided background on the current state of some of the major federal household surveys in the United States and outside perspectives on how other nations handle many similar difficulties in household data collection. The first talk in this session focused on a review of the current U.S. federal household data collection system. Subsequent talks presented foreign case studies: the current United Kingdom (U.K.) model for survey integration; the case of the Netherlands, which relies less on household surveys and more on official population registers; and Canada's use of a multipronged approach to improve efficiencies, including establishing a corporate business architecture and developing a strategy of survey integration. The international examples of survey data collection served to open up a broader discussion about data collection approaches to consider.

FEDERAL HOUSEHOLD DATA COLLECTIONS
IN THE UNITED STATES

Katharine Abraham (University of Maryland) highlighted three major aspects of the federal statistical system: (1) the current survey environment is difficult, (2) data users have become more demanding of survey data, and (3) the system is searching for solutions. Specifically, she described several data collection challenges that have contributed to making the current survey environment increasingly difficult. One of these issues is the quality of survey frames. Survey practitioners and researchers agree that, generally, household survey frames provide poor coverage of several important segments of the population.

5

Another issue is that it has become increasingly difficult to reach respondents. It is also increasingly difficult, once people are reached, to convince them to grant an interview. Finally, increasing concerns about privacy and confidentiality have exacted a toll on survey participation.

Coverage patterns in many federal household surveys are evidence that survey frames are not always adequate to reach a representative sample of the population. As Abraham noted, coverage ratios for personal visit surveys tend to be lower for black respondents than for nonblack ones; they are lower for men than women; and they vary systematically by age. Despite coverage ratios that generally trended downward from 2000 to 2008, coverage ratios for the American Community Survey (ACS) have, by contrast, been higher and more stable than those of other Census Bureau surveys. To help combat the coverage problem, the Census Bureau, in its 2010 survey redesign process, decided to use the continually updated Master Address File (MAF)—the frame the ACS uses—as the frame for its other current surveys. The use of the MAF will begin with the 2014 surveys.

Another problem creating challenges in the survey environment is the increasing difficulty of contact with survey respondents. Gated communities restrict access to respondents for in-person interviews and nonresponse follow-up. The use of voicemail and caller ID helps respondents avoid contact with an interviewer in telephone surveys: they can let calls go to voicemail or not answer calls from numbers they do not recognize on their caller ID display. The number of cell-phone-only households has risen sharply in the past 10 years and continues on an upward trend, thus making an initial contact through a telephone frame more difficult in the case of these households.

Obtaining respondent cooperation has become increasingly difficult. Abraham explained that increasing demands on respondents' time, such as long commute times and increasing numbers of telephone solicitations, make respondents less likely to cooperate with an interview request. Furthermore, survey requests, such as from the federal government, compete with multiple other surveys and sales solicitations for the already limited time and interest of potential respondents. Finally, pervasive concerns about privacy and confidentiality among many in U.S. society hinder survey participation. It is not only the federal government and its data collection contractors that suffer from an increasingly unfriendly and costly climate for surveys; other survey research organizations are also encountering similar problems.

In addition to an increasing unwillingness to participate in surveys, there is also evidence of rising item nonresponse within surveys. As an example, Abraham cited a study by Bollinger and Hirsch (2006) showing that item non-response has increased on the Current Population Survey's usual weekly earnings question. Increased item nonresponse is further evidenced by increasing imputation rates on questions of wages and salaries. By 2000-2004, imputation rates for weekly earnings were up to about 30 percent for survey respondents.

Next, Abraham briefly discussed the increasing demands from increasingly more sophisticated data users. Data users tend to demand more timely and comprehensive data. Many have pushed for more detailed data—that is, data on small geographic areas and population subgroups. There has also been a call in the data user community for better integration of estimates (e.g., income, disability, poverty) from different sources.

Agencies have used multiple strategies to increase or maintain current survey response rates. Some surveys use advance notification mail materials or offer multiple modes for response. Other means used are increasing the number of contact attempts with respondents, improving interviewer training, and, in the case of the ACS, making the survey mandatory. Some surveys offer incentives for participation. Abraham noted, however, that the evidence of the effectiveness of any of these methods is limited, and their use comes with increased survey costs.

In addition to these strategies, Abraham laid out possible actions that agencies could take to meet the challenges facing federal household surveys. Although the last two years have seen an increase in funding for some statistical agencies, it is unlikely that increases will continue, particularly in the current political climate with calls for reduced government spending—making it even more important to look for ways of increasing efficiencies.

Frame improvement is one area in which agencies are attempting to identify opportunities for increased efficiency. As mentioned earlier, the Census Bureau will begin using the MAF for many of its personal visit surveys. In addition, the ACS will be used to provide stratifications for sample designs by providing more current information on the characteristics of geographic areas. Abraham asked if, in addition to this change, the ACS should be used directly as a sample frame itself.

Other frame improvement ideas include incorporating cell phone numbers into random digit dialing (RDD) samples. The use of the Internet for survey administration would be most cost-effective; however, there is not yet any agreed-on methodology for creating a frame for online surveys. While online surveys remain an attractive prospect for survey administrations, Abraham stated that more work is needed on how the web option can be most effectively presented and on ensuring web-reporting data quality.

Administrative records are another avenue agencies are pursuing for use as sampling frames, as survey benchmarks, as sources of auxiliary data for model-based estimates, and for direct analysis. This is a promising area for future research, Abraham said, but she added a word of caution about treating administrative records as the "gold standard" of data, because little is known of their error properties.

Better methodologies could be explored for use to reduce nonresponse and imputation rates. For example, paradata (i.e., data automatically generated by electronic data collection tools about the survey process) and better survey

frames could aid in improving nonresponse adjustment. Of particular interest is the potential role of the ACS, or some other large data set, as a sampling frame. This could provide better information on both respondents and nonrespondents—information that could be used for better adjustments.

Model-based estimates are another methodology to make greater use of. These have become increasingly accepted as a viable alternative to direct estimates, particularly as direct estimates for small areas become prohibitively expensive. The ACS is important here, too, in that it may be a valuable source of auxiliary information for use in small-domain models.

Outside the technical aspects of federal household surveys, it is worth considering the organizational environment in which these surveys are conducted. Improved interagency cooperation and coordination are essential. For example, the Census Bureau could facilitate this by more transparent cost accounting for client agencies, giving agencies greater input on infrastructure decisions that affect their surveys, as well as giving them broader access to frames and survey data that are important to accomplishing agency missions. Title 13 of the U.S. Code (the law that guarantees the confidentiality of census information) is a factor that must always be considered with respect to who gets access to what data. Yet it would be extremely valuable to client agencies to have access to the sampling frames used for their surveys and to have more access to the information that is collected, particularly if an agency wished to go back to a set of respondents.

Clearly, federal statistical agencies face an increasingly difficult environment for collecting data as well as growing demands with respect to the data that are collected. A substantial amount of research is being done to meet these challenges, but strong interagency collaboration is going to be critical to efficiently implement the new ideas coming out of this research.

SURVEY HARMONIZATION IN THE UNITED KINGDOM

Cynthia Clark (National Agricultural Statistics Service) presented an overview of the U.K.'s approach to household survey harmonization in government surveys. Paul Smith from the U.K. Office of National Statistics (ONS), the author of the presentation, and one of the prime contributors to the work on the U.K. Integrated Household Survey (IHS), was not able to attend the workshop. Clark explained that the focus of the presentation is on the original design of the IHS but includes a discussion of the challenges the United Kingdom has faced related to the design over the years.

Responding to many of the same pressures that confront household surveys in the United States and as part of the U.K.'s survey modernization program, the ONS developed an Integrated Household Survey design. The basic concept was to develop a framework in which multiple household surveys could be integrated into a common design. In the United Kingdom, household surveys

have developed independently, much like in the United States. Each had different objectives and different methodologies for obtaining the ideal survey sample for a given topic area. For example, the Labour Force Survey (LFS) is not a clustered design, whereas many of the ONS's other household surveys are clustered. The integrated design increases the sample size for core variables by asking them on all the component surveys.

The design of the IHS relies on the use of modules formed from four existing continuous household surveys: the LFS (including some regional supplementary surveys), which serves as the IHS survey core and provides the majority of sample cases (200,000 households); the General Lifestyle Survey (formerly the General Household Survey); the Living Cost and Food Survey (formerly the Expenditure and Food Survey); and the Opinions Survey (formerly the Omnibus Survey). After the original modular design incorporated these four surveys, others, such as the English Household Survey, were added. The idea behind the modules was to standardize concepts and questions across the surveys. In its current form, the survey sample includes 265,000 households and uses a staged approach.

Figure 2-1 shows the modular structure of the surveys. The vertical axis on the graph represents the sample cases, and the horizontal axis the different modules and interview length. All interviews include the core survey, followed by a rotating core. The remaining modules represent different surveys presented to different respondents. Parts of the sample are visited quarterly over five quarters, parts are visited annually over four years, and parts are visited only once.

Such an undertaking, Clark noted, relies on several critical assumptions about changes. First, the flexibility of the field staff must be increased, and interviewers have to be trained to do all interview types. Surveys with an original clustered design are ideally unclustered to be joined with the core LFS, which has benefits in reduced variance of estimates.[1] Content and procedures require standardization among the surveys. Finally, increases in sample size for core variables help to improve small-area estimation.

The expected benefits include reduced sampling variance due to increased sample size, cost savings associated with the unclustering of the sample designs, and two-phase calibration, which will enable the use of the estimates from the core in calibration for components. The increased sample size of the core is expected to produce a variance reduction of up to 20 percent for the LFS (if fully unclustered). An unclustered design for the non-LFS surveys is expected to reduce variance of the module variables by 2-15 percent, although this has not yet been implemented.

One of the many challenges encountered was the implementation of the IHS in the field. Originally, an entirely new case management system was

---

[1]The unclustered design would sample addresses directly from the Postcode Address File (PAF) rather than selecting them from a subset of postal code sectors (Office for National Statistics, 2010).
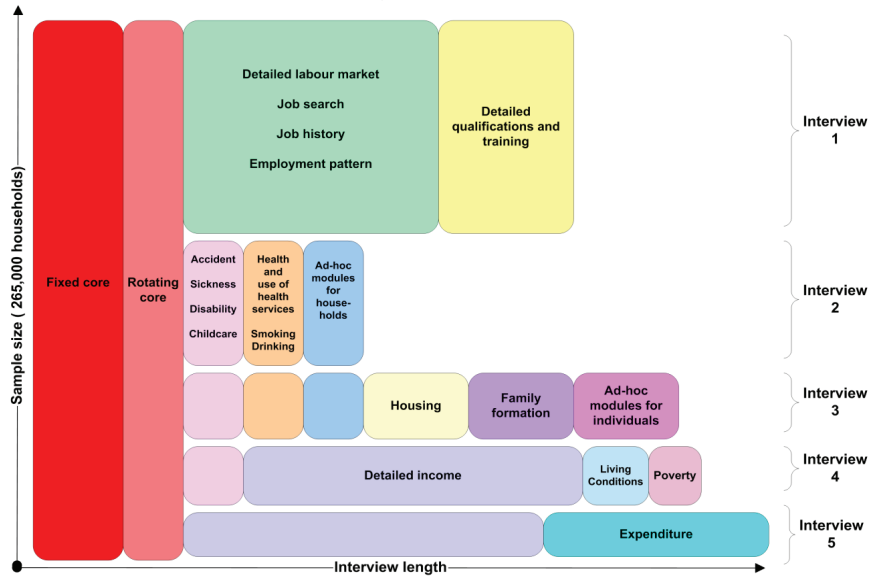
FIGURE 2-1 Illustrative diagram of a modular continuous population survey.
SOURCE: Workshop presentation by Cynthia Clark based on Office for National Statistics public sector information licensed under the U.K. Open Government Licence v1.0.

planned as part of the field office modernization for the IHS, but the office modernization project turned out to be too ambitious. Instead, field operations had to fall back on existing survey systems. Given that data users do not like to see variables dropped, another problem was that the survey core ended up being too long to be practically administered in the field. Problems related to inconsistencies in the survey outputs also persist. The two-phase calibration has only been partly implemented so far. The calibration works, building in automatic consistency, which increases the quality and usability of outputs, but it has shown only marginal variance gains. Estimates from the IHS are currently released as "experimental," which allows data user input and feedback to quality-check the procedures and outputs; they are not yet classified as "national."

Although the implementation of the original design has proven to be challenging, many of the difficulties were due to the necessary systems not being in place. Stepping back has made survey harmonization both more important and more challenging. Despite the difficulties, there has been considerable progress in the design and implementation of the IHS.

DISCUSSION

Hal Stern invited the workshop attendees to ask questions of the first two presenters. Phillip Kott (Research Triangle Institute) directed the first question to Clark: what did the author from the ONS, Paul Smith, mean by the unclustering of the current LFS in the United Kingdom, and how would this save money? Clark explained that the LFS was already unclustered, and, since it was the largest of the surveys, it made sense to move the smaller surveys to that design. Because Clark was not the author of the presentation, she referred to Paul Smith's paper for additional information about the plans related to unclustering (Smith, 2009).

Eric Bergman (Bureau of Labor Statistics) noted that there are certain economies of scale to combining these surveys and asked whether there were any initiatives to make the IHS mandatory. Clark responded that there were no initiatives along those lines.

Lawrence Brown (University of Pennsylvania) asked how the integrated survey design affected the longitudinal character of the LFS and how this would be reflected in the other integrated surveys. Clark said that she did not have enough information about the design of the other surveys or if they had longitudinal components in them, but the LFS in its current form is conducted in 5 segments over the course of 15 months. Stern wanted to better understand how modules moving into and out of the integrated survey would look over time and if there are forecasts regarding ultimate costs for the IHS on a large scale. Clark said she did not have an answer to those questions.

Abraham asked about the total time required to administer the survey. Given the length of many of the surveys in the United States, it would be difficult to see how this model could be applicable here, she said. Clark noted that the LFS core of the IHS is approximately 20 minutes, and some of the other modules rotate in and out.

Robert Groves (Census Bureau) made the point that there is nothing inherent in the design of the IHS to say that questionnaire length could not be constant across interviews, through appropriate matrix sampling of the modules. Furthermore, although the ONS is not doing this, administrative records could be used to guide inclusion probabilities for the matrix sampling. In other words, there would be an administrative data-driven inclusion probability for rotating modules.

Andrew White (National Center for Education Statistics) asked whether the push for integration was budgetary in nature. He also asked whether the United Kingdom has been experiencing challenges related to household survey data collections similar to those in the United States and whether the ONS expects the harmonization to address these problems. Clark responded that funding became available for infrastructure development, which represented an incentive to embark on this project. The primary reasons for doing this were

not necessarily in response to the types of challenges described by Abraham in connection to the U.S. household surveys, she said.

Katherine Wallman (Office of Management and Budget) commented that it appears that the IHS was not designed with the goal of reducing response burden. When the U.S. Government Accountability Office (GAO) has prepared reports on the federal household survey network in the past, its perceptions were that the surveys are duplicative and a heavy burden on respondents. The GAO wanted to know why surveys are not combined together in a framework similar to the IHS, but it appears that the IHS has grown out of different considerations. It is also interesting to hear that some of the supplements are included only periodically.

Graham Kalton (Westat) asked how difficult it was to bring together the existing surveys and whether there was any infighting, given response burden constraints and the probability that the sponsors of each of the existing surveys had different interests and agendas. Clark said that in her experience this was not a major problem. There was a significant push for harmonization and modernization as part of the integration process, which may have facilitated their willingness to compromise. However, she added, the integration process has not completely succeeded yet, and the LFS still publishes its own estimates, rather than the IHS estimates.

Alan Zaslavsky (Harvard Medical School), asking what an acceptable "national statistic" entails, said that there are several potential problems related to generating such a statistic. One issue might be the technical and operational quality of the systems used to generate the statistic and whether they are working correctly and are doing, procedurally, what they are supposed to do. Another issue might be the acceptability of the estimation methods, as these become more complicated than simply asking 1,000 people a question and tabulating the numbers. He asked about the importance of these considerations as the new methodology is implemented in the United Kingdom and whether acceptance has been built for these new methods of estimation.

Clark responded that, in her opinion, an important part of the transition to a national statistic is ensuring that the data remain relevant when compared with past data and specifically ensuring that there are some mechanisms for benchmarking and for helping users to understand the new data series. The ONS uses quality measures similar to those used in the United States—timeliness, accessibility, comparability, accuracy, relevance, and consistency—in determining what will become a national statistic. If, in fact, new estimates are adequately bridged to previous data, then, generally, after several years, a statistic will move from an experimental one to a national one. Small-area estimation procedures were also used for the first time in official statistics, after a period of being considered experimental.

Barbara O'Hare (Census Bureau) asked how the federal statistical community can move toward greater acceptance of model-based estimates, similar

to what was done with small-area estimates in the United Kingdom. Clark suggested that the U.K. model of labeling model-based estimates as experimental until they have gained acceptance (and can become national statistics) could be a model for the United States as well. The Small Area Income and Poverty Estimates (SAIPE) Program at the Census Bureau is an example of publishing model-based estimates in the United States, but when these data were first released, not all users were comfortable with using them. The estimates were released because they were better than anything else available, and they were labeled to advise data users to exercise caution when using them. However, this does not always help in gaining acceptance.

Constance Citro (Committee on National Statistics) noted that the SAIPE estimates are available and are being used, although it is not wise to spring new data on users overnight. It is imperative that the statistical community have a dialogue with data users and describe the positives of model-based estimates, such as stability over time. Once they understand what they are dealing with, they will want the data.

Returning to the topic of challenges related to nonresponse, Jelke Bethlehem (Statistics Netherlands) commented that, on the basis of his 30 years of experience working on the issue of nonresponse, he now thinks that the focus should be on the composition of the responses, rather than on trying to improve the response rates. If an organization spends enough time and money, it is possible to increase the response rate, but research shows that this sometimes makes the responses less representative of the sample. Instead, the focus should be on measures that help balance the response.

Abraham agreed that increasing response rates at all costs should not be the objective, but she expressed concern about measures taken to balance a sample. In some cases, balancing the sample along demographic variables works well, but there may be other variables of interest for which it does not work. She noted that the approach of balancing the sample sounds similar to a quota sample, and experience shows that quota samples do not perform well, at least in the case of establishment surveys. Clark added that one of the objectives of an adaptive design of this type is to enable researchers to evaluate the composition of the respondents, and that it helps to have paradata to be able to monitor the sample in real time.

Citro made the point that great design ideas alone will not solve the current problems of the federal household surveys. The success of integration depends at least as much on systems, procedures, and cost accounting as it does on design ideas. She referred to Clark's discussion of the problems with the case management system, which were a problem with the 2010 census as well. The question—and challenge—for the statistical agencies is to work together to do better than in the past in improving the basic components of the survey "manufacturing process."

STATISTICS WITHOUT SURVEYS?
DATA COLLECTION IN THE NETHERLANDS

Continuing the focus on foreign survey systems, Jelke Bethlehem (Statistics Netherlands) presented an overview of the way the Dutch statistical system collects national data and discussed the population register that serves as a backbone to an integrated information system. He began by walking the audience through a brief history of the census and survey systems in the Netherlands.

The Netherlands has a mandatory national register, which has been digital since 1994. It no longer fields a census in the traditional sense, instead conducting a virtual census, which involves information gathered from the population register and through surveys. Demographic data are obtained from the register, and socioeconomic data are gathered via the LFS.

Statistics Netherlands successfully uses the population register for three main applications: (1) as a simple and quick data source for monthly population statistics (only counts, not estimates), (2) as a sampling frame for surveys (for persons only, households must be constructed), and (3) as a source of auxiliary variables for weighting adjustments to correct for nonresponse.

Responding to increasing calls for more comprehensive, higher quality data, Statistics Netherlands created the Social Statistical Database (SSD), an amalgam of the population register, the LFS, the Survey on Unemployment and Earnings, and other administrative sources. In the case of the Netherlands' 2001 census, the SSD was used with much success to meet the European Union's demand for greater census detail. Using the SSD, the work of putting together a census was completed early, despite getting a late start, and at a cost of €3 million, versus the €300 million a traditional census would have cost. SSD data can also be linked to both survey respondents and nonrespondents.

Despite the reliance on the SSD, Bethlehem said, there is still pressure to reduce response burden. As a result of this pressure and budget constraints, the focus of data gathering has shifted to more secondary data collection, mostly from registers. In this context, Bethlehem mentioned the Netherlands Statistics Law of 2003, which stipulates that surveys should occur only when the data are not available elsewhere. It also gives Statistics Netherlands access to all government registers.

Naturally, the population register is not error-free, and some of the data require substantial editing. One of the main reasons for the errors relates to students who tend to move and not register. The fact that Statistics Netherlands does not control the data collection is also a challenge because of a lack of understanding of quality control and definitional problems, he noted.

The government can mandate changes in the registry data at any time, a circumstance that can also lead to problems. The data for the construction sector are an example of this; the sector reports its earnings via tax administration. During a recent economic crisis, companies were allowed to change their

declarations from monthly to quarterly. This introduced a lack of comparability and problems with the reliability in economic data in the construction sector.

To keep pace with increasing data demands and shrinking budgets and to combat current data collection problems, new ways to collect data are under study, Bethlehem reported. One strategy is to collect data directly from the administrative and financial systems of companies. Another is to use radio frequency identification tags (RFID) and global positioning systems (GPS) to collect transportation statistics. The use of online robots that collect data from specific websites allows for the leveraging of information already available on the Internet. One possible use of such a robot is for the collection of price data to produce a consumer price index. Of course, he said, there are many questions surrounding these data collection methodologies. Do they work? Are they legal?

Bethlehem concluded by saying that, despite opportunities for using registers and technological advances for data collection, there will still be a need for surveys in the future. It is likely that the surveys of the future will be increasingly Internet-based or mixed-mode, although these present new challenges, such as mode and selection effects, that are difficult to separate. There are other methodologies yet to be considered, and Statistics Netherlands is keeping an open mind about the possibilities.

CANADA'S HOUSEHOLD SURVEY STRATEGY

Jean-Louis Tambay (Statistics Canada) presented another perspective from outside the United States, by giving an overview of the Canadian household survey system. Table 2-1 lists major Canadian surveys with monthly data collection. Currently, Statistics Canada has three major sampling vehicles for household surveys: (1) the LFS area frame design, (2) RDD, and (3) a census of population, conducted every five years. Many household surveys draw their samples from LFS sample clusters, are administered as supplements to the LFS questionnaire, or, to cover certain population subgroups, survey recently rotated-out LFS sample units. Like other nations, Canada faces an increasing demand for survey data—a demand that exceeds the current capacity of the LFS to provide samples. New solutions are being proposed and tested to address the limits of the current survey platform, which involve the flexibility and timeliness of surveys (especially developing computer applications for surveys), costs, response burden (particularly for LFS respondents), falling response rates, coverage problems with RDD and telephone surveys, and the challenges of surveying difficult-to-reach populations.

In response to the demand for data, Tambay said, Statistics Canada has developed several strategies grouped under the term "New Household Survey Strategy," including survey integration, spreading interviewer and response burden, development of a master sample, creation of a population frame, and

TABLE 2-1 Major Canadian Surveys with Monthly Data Collection

Survey	Size	Details
Labour Force Survey	60,000 households (120,000/year)	6-month rotation (10,000 new cases/month); telephone-first contact for 36% of new cases; use Address Register to replace/supplement listing activities
Canadian Community Health Survey	65,000/year	50% CAPI (LFS area frame); 50% CATI (telephone lists); pool 2 years' sample for small health regions
Survey of Household Spending	20,000/year	LFS area frame
General Social Survey	25,000/year	Random digit dialing
Travel Survey of Residents of Canada	110,000/year	LFS "live" supplement
Canadian Tobacco Use Monitoring Survey	50,000 households/year 20,000 persons/year	Random digit dialing

NOTE: CAPI = computer-assisted personal interviewing, CATI = computer-assisted telephone interviewing, LFS = Labour Force Survey.
SOURCE: Workshop presentation by Jean-Louis Tambay.

integration of listing activities. The process of survey integration includes using a common core of questions for all surveys, harmonizing content modules, creation of a master sample, and integrating survey and census listing activities. Spreading interviewer and response burden was achieved, in one case, by spreading the collection period for the Survey of Household Spending over a 12-month period, rather than the 3-month collection period that was used in the past. The Canadian Community Health Survey (CCHS) sample of 130,000 respondents was divided in half, and data collection was spread over two years, instead of using the whole sample every other year. Finally, Statistics Canada is considering ways to increase response options, such as offering electronic data reporting, which is currently used for business surveys and was also tested during the 2006 census.

Of the four options considered for the design of a master sample, it was decided to create the sample by pooling first-phase surveys but to limit the surveys used to just the LFS and the CCHS. The sample was created, and a pilot survey was conducted in 2008 using an existing survey vehicle, the General Social Survey (GSS). Tambay said that this was complex to implement because it was difficult to develop the proper weights and variances.

Furthermore, there had to be a way to deal with samples that were not really independent. The results were disappointing: response rates were low and design effects were high. The master sample option was thus abandoned, and the idea of using the census as a frame was reopened. A population frame (of persons) created from census follow-up was considered in lieu of the master sample design, although this type of frame also suffered from problems, particularly privacy concerns.

Integration of listing activities involves the coordination of census and the LFS cluster listing activities via a common listing application. To aid in cluster listing operations, Statistics Canada provides its interviewers with dwelling lists from the Address Register (AR), which is similar to the U.S Census Bureau's Master Address File. Used since the late 1980s, the AR is derived from telephone billing files from many major telephone companies and Infodirect (similar to a white pages compilation of all Canada), plus other smaller sources, such as tax rebate records for new dwellings.

The AR was used to define mailout areas for the 2006 and 2011 censuses, which account for 70-80 percent of the country. In 2004, it was also used to replace or supplement the LFS listing in many clusters. For the 2011 census, a continuous listing was introduced to update the AR (for the 2006 census, the AR was updated through a full-scale block-canvassing operation that took place the previous fall). Leading up to the census, interviewers would verify only clusters that AR methodologists believed were in substantial need of updating, with the assumption that about a third of the clusters would be visited for continuous listing. This is what gave rise to the idea that if interviewers were in the field to do listing for the census, the activity could be combined with listing for the LFS, Tambay noted. The LFS usually conducts its own listing activities, although for about 40 percent of the clusters, the AR is considered of good enough quality to dispense with the initial listing. In another 20 percent of the LFS clusters, AR dwelling lists are updated by interviewers, and in the remaining 40 percent AR coverage is such that it is deemed preferable to have interviewers develop new dwelling lists.

Tambay explained that the process for integrating survey and census listing activities had three components: (1) coordination of census and the LFS listing activities, (2) development of a common listing application, and (3) increased use of the AR to replace or supplement the LFS listing. The coordination component consisted of positive and negative coordination. Positive coordination meant that if a cluster for the LFS has to be listed in a certain month and the AR has to list it sometime before the next census, then Statistics Canada tries to coordinate the process so that the cluster is listed for the AR before it is needed for the LFS. Negative coordination means the listing for the AR is skipped for clusters in which LFS is actively interviewing.

The latest innovation at Statistics Canada is a corporate business architecture, Tambay said. The goals are to be more efficient, robust, and better able

to respond to new developments. Two of the main principles are (1) decision making optimized across the organization and (2) centralization of such processes as staff services or information technology services and infrastructure.

Several proposals for social surveys have come out of the new program, including creating a household survey frame function and developing a social survey processing environment that is common to multiple surveys as well as increasing the use of electronic data reporting. The LFS is ideal for testing electronic data reporting because survey respondents have the option of providing an email address in their first month in the sample or responding for the following five months of the survey via an Internet address provided by Statistics Canada.

To address the first proposal, the household survey frame project was created. One activity for this project is to improve AR quality and content. This means it is necessary to increase the availability of phone numbers, maximize AR coverage, and increase AR content. The plan is to achieve this through several steps. First is to increase the availability of phone numbers, which mostly come from billing files and Infodirect. Phone numbers are then supplemented with information from the census or tax data. However, the 2006 census did not provide much more information than Infodirect already had. Telephone numbers from tax files are also problematic because the number could be for an accountant who prepared the return or a work number. The child tax benefit file has proved to be a more useful source of telephone numbers, and it tends to cover households with young children, Tambay observed.

Other indirect methods of obtaining more complete information are also under consideration, such as matching tax records to Infodirect phone numbers to add apartment numbers that are missing on Infodirect. Exploring a cell phone billing file was also attempted. An application to sample from this frame has yet to be developed. A consequence of trying to add additional phone numbers to the frame is that regional offices are communicating that their telephone centers are already operating at capacity with the phone numbers that are currently in certain frames.

Statistics Canada is also attempting to expand its address resources using such tools as municipal lists and tax forms. Frame coverage in the AR currently is 96-97 percent, with 85 percent of these addresses being mailable. In addition, the Canada Post Corporation Point-of-Call file, which is comparable to the U.S. Postal Service's Delivery Sequence File, is also a very reliable source, especially in urban areas.

Another goal of this activity is to improve AR content by creating a person frame. The census short form, which has household composition information, and the tax family file, which is a file that is constructed from tax records, can be used to construct this frame. Because people tend to declare their children, coverage is about 96 percent. That will be used to update the census information.

The second activity of the Household Survey Frame Project is to develop a common frame for household surveys. This would entail establishing processes for sample management (to control respondent burden), completing integration of the AR with the LFS area frame, and developing a methodology for the use of phone numbers in the design of computer-assisted telephone interviewing (CATI) surveys.

There are several keys to a more complete integration of the AR with the LFS, Tambay noted. The first is two-way communication on new dwellings. If any growth is identified through the LFS or the AR, then one should be communicated to the other to get the best possible integrated address. The second key is an ongoing attempt to integrate into the AR noncity-style addresses, for example, postal installation addresses consisting of a type of delivery, which may be general delivery; lock box number; or municipality name, province, and postal code. Finally, every attempt is being made to identify AR needs for the 2014 LFS redesign.

Although still in the planning stages, researchers are currently attempting to develop a methodology for the use of phone numbers from the improved frame in the design of CATI surveys. The goal is to pilot this methodology on the General Social Survey in fall 2011.

For the future, Tambay said, the next thing to consider may be sample coordination (rather than coordination only for frames). Tied to the LFS redesign is the redevelopment of the generalized sampling system. Statistics Canada would also like to develop a new system for selecting dwellings. For the portion of the LFS that can utilize the AR, options for keeping this frame current include updating it by administrative sources and forgoing listing, taking simple random samples of subclusters, and sample coordination with other surveys to avoid visiting the same respondent too often.

DISCUSSION

Chester Bowie (National Opinion Research Center), session discussant, observed that one of the themes of the morning's session that sets the context for the rest of the workshop is that surveys have become more complex and difficult over the past 10-15 years. A number of factors drive this complexity: quality and cost concerns related to sampling frames, increasing nonresponse rates; privacy and confidentiality concerns; and rising survey costs, with concurrently shrinking budgets. The statistical community is also not yet sure how to best use administrative data or model-based estimates. Each of the countries represented at the workshop is addressing these issues differently.

The United Kingdom has standardized and integrated its major household surveys. This is an intriguing idea, Bowie said, but such a system would be much more difficult to implement in the United States, where the statistical system is more decentralized. Several past attempts to standardize basic

demographic questions across surveys at the Census Bureau were unsuccessful because each survey sponsor had its reasons for wanting to ask a specific question in a particular way.

The Netherlands Social Statistical Database is interesting because it is a move away from surveys toward population registers, Bowie said. This lowers survey costs, but there are issues inherent in gathering data this way. Canada has addressed some of its challenges through the use of master sample frames and samples, integrated listing activities, and household survey sample coordination. Some of these strategies are unique.

Some have argued that the current approach to conducting household surveys in the United States is unsustainable. Bowie reiterated that this problem is the focus of the workshop and that serious thought should be given to what can be done in the future to address it.

Hermann Habermann (Committee on National Statistics) sought clarification on the use of population registers in the Netherlands. If it was a distrust of government that made people wary of censuses, how was a register received? A register can be perceived as even more pernicious than a census. Bethlehem said that there has always been a good population register in the Netherlands. This became an issue during World War II because religion was recorded on the register and, when the Germans invaded, they were able to easily identify Jews in the country using the register. Today, there is a variety of registers, and they seem to not bother people anymore. Many, if not most, people in the United States may be in registers without even knowing it.

A follow-up to Habermann's question concerned the political discussion on using registers instead of surveys in the Netherlands. Were privacy advocates concerned that the combination of registers would be a threat to privacy? Bethlehem responded that the only political discussion was about reducing the administrative burden of government. No privacy issues were raised when the bill was proposed in Parliament, and the public really does not seem to be concerned about it.

Wallman asked if registering was mandatory in the Netherlands, as it is in Germany. She wondered whether there would be an adverse reaction to such a requirement in the United Kingdom or the United States. Bethlehem again noted that most people in the Netherlands probably do not even realize that they are in the population register. The only time citizens encounter the register is when they have to renew a passport or when they move and they are required to fill out a form on the Internet. In situations like that, it can become a problem if they are *not* in the register. However, the fact that the register is mandatory has never surfaced as an issue.

Tambay recalled a case in which a journalist discovered that the department that administers unemployment benefits in Canada has been maintaining a data file on the labor force. The Canadian government publishes what files are used by which government departments every year, so the existence of this file was

always public information. Yet, when the journalist brought attention to this, a scandal followed that affected subsequent data collection efforts, because fewer people were willing to share information with this particular department after the incident. The department was also ordered to destroy the file, because although the existence of the file was always public, information about how the data were being used was found to be not transparent enough.

Robert Kominski (Census Bureau) suggested that a synthetic register, or one compiled from several data sources, may be a viable concept in the United States. There are already many data systems here, and these could be used to develop an effective register. An example of an existing register in the private sector is the charge card registration system, which includes point-of-purchase data and other information. The banks are authorized by the federal government to collect these data, and the federal government could say that these data are within its purview. Kominski added that perhaps this is a radical idea, but the purpose of the workshop is to think broadly.

He went on to say that, in the current political climate, U.S. residents might be willing to give up their privacy and register, if they thought that such a system would prevent public services from being delivered to those who "do not deserve them." Some people might do this to obtain greater security or, in their eyes, fair administration of state and federal goods and services. Some might be offended by these ideas, he said, but there is a very large segment of the population that would not be.

A workshop participant noted that even if only 5 percent of the population refused to get an identity card or register, that is still 5 percent of the population that would be missing, which would ordinarily be considered unacceptable.

Wallman did not think the issues surrounding registers were necessarily related to whether or not the registration was mandatory, but rather, in talking to colleagues in other countries, whether or not the register was tied to certain benefits. For example, eligibility for child care in the Netherlands is entirely tied to the registration of that child. Such a setup would have a huge impact here. There may be pros in addition to the cons typically associated with registers, she said.

Lawrence Brown cited the example of Israel, which has a census as well as a registration system. Although this system is far from perfect, particularly for households, the government is building a secondary system of dual-system estimation to correct the registry lists for census purposes. A question that remains, however, is how a system like this can be built into a household data system with the same effectiveness. Another question pertains to inaccuracies in the registration system. Although the register in the Netherlands enables a count of the population, there do not seem to be good address records. He asked: Would it be better to have a dual-system follow-up to correct these inaccuracies?

Bethlehem said there were about 2,000 persons in the Netherlands not in

the national register, and they are most likely illegal immigrants. About 15 percent of the register records contain errors, but these errors come from incorrect addresses. If someone is listed at an incorrect address in the register, this can become a problem for them should they wish to, for example, get a passport. Because people depend on the register to receive services, it tends to be fairly accurate. Statistics Netherlands defines survey populations to be the population in the register, thus that sampling frame completely fits the population. There is also a database for information to do weighting adjustments. The question of whether including illegal immigrants in the count and surveys is a problem is a decision each country has to make.

3

Sampling Frames

Graham Kalton, moderator for the session, described the presentations as a discussion of the potential uses of sampling frames to aid in particular surveys and the multiple sources for these frames. Given the costs associated with frame development, some of the questions to consider are whether there are any economies that can be achieved with the current sampling frames and what are the difficulties related to implementing them.

USING LARGE SURVEYS TO ASSIST IN FRAME
DEVELOPMENT FOR SMALLER SURVEYS

James Lepkowski (University of Michigan) began his talk on using large surveys as frames for smaller surveys with examples of cases in which this is currently being done and a discussion of the issues associated with these approaches. The first example described the Current Population Survey (CPS) and the American Time Use Survey (ATUS).

The CPS is a well-established, rotating panel, continuous survey of the noninstitutionalized population in the United States ages 15 and older. A joint effort of the Census Bureau and the Bureau of Labor Statistics, the CPS is the primary source of information about characteristics of the U.S. labor force. It uses independent samples in each state and the District of Columbia and oversamples the Hispanic population. Since the 1940s, it has used probability sampling and has produced national as well as state-level estimates.

The ATUS uses a sample of households from a CPS panel that is rotating out of the survey. There are three stages of the ATUS sample design. From

23

the sample of households (in the third and final stage of the sample design), one person age 15 or older is randomly selected for interview by telephone and becomes the ATUS "designated person." Nontelephone households are contacted by mail, given a phone number, and requested to call in, with a $40 incentive that is awarded at the completion of the survey.

Lepkowski said that one of the major challenges in using the CPS as a frame for the ATUS is timing. Although most of the CPS sample becomes available to the ATUS within three months, the sample is still spread out over time due to the interviewing and processing schedule. Similar challenges related to timing have led some survey organizations to abandon second-phase samples.

Another challenge in the context of the CPS and the ATUS is that the CPS is a household survey, which must then be transformed into a person-level sample for the ATUS. Finally, it is possible that ATUS response rates are adversely affected by previous participation in several prior CPS interviews, but it is difficult to determine conclusively the potential magnitude of this effect. Overall, the telephone response rates are in the mid-50 percent range.

The second example Lepkowski described is the case of the National Health Interview Survey (NHIS) and the Medical Expenditure Panel Survey (MEPS). The NHIS is the primary source of data about the U.S. household population's health and health care utilization. The survey is conducted by the Census Bureau and sponsored by the National Center for Health Statistics (NCHS), although other agencies also fund supplements, a situation that can be an important factor that influences an organization's ability to share sample efficiently. The NHIS is a continuous, multistage, national probability survey with oversamples of black, Hispanic, and Asian populations. Response rates vary depending on the type of interview, generally ranging between 65 and 80 percent.

The MEPS, sponsored by the Agency for Healthcare Research and Quality (AHRQ), uses completed NHIS interviews as a sampling frame for the household component of the survey (there is also a medical provider component and an insurance component). The goal of the survey is to produce national and regional estimates of health care utilization and expenditures. Approximately 15,000 households are included annually, with occasional oversamples for additional policy-relevant subgroups. The MEPS also utilizes the oversampling performed for the NHIS. Rather than a cross-sectional design like the NHIS, the MEPS uses a panel design.

The MEPS response rates are also affected by the response rates to the NHIS. Response rates for recent NHIS surveys have typically been in the upper 80s, and the MEPS nonresponse rate is compounded by the nonresponse in the first phase. In addition, the NHIS sample sizes can vary from year to year, changing the proportion of the sample the MEPS takes from the NHIS to meet its own sample size designations.

One of the main advantages of using one survey as the sampling frame

for another is the cost efficiency that can be achieved by the second survey. The cost savings can be realized in the form of efficiencies in sample design, data collection, screening, and data processing. For example, the ATUS has a list of items that are nearly identical to those in the CPS, and going through the same processing system saves the cost of system development. Although typically the efficiencies benefit the second survey, Lepkowski observed that when the sample sharing is a long-term arrangement, there has to be some sharing of the cost burden as well.

He pointed it out that there are several challenges related to these designs as well. Nonresponse rates can be affected not only by the fact that respondents' willingness to participate sometimes declines by the time of the second-phase survey, but also because of increased difficulties related to locating sample persons by the time of the follow-up. Although drawing a sample based on another survey also presents a unique opportunity to estimate nonresponse bias based on responses to the first survey, this is often leveraged to some extent, but perhaps not as much as it could be. A related concern is the measurement bias that can potentially be introduced into the second-phase survey as a result of participation in previous surveys, even if respondents are willing to participate (also known as time-in-panel bias).

The quality of any stratification performed for the second-phase survey depends on the quality of the data collected in the first survey. For example, if the second-phase survey is stratified on income and this information is mis-reported in the first survey, the misclassification will lead to inefficiencies in selection.

Capacity issues are often another consideration. The first survey has to provide adequate sample to meet the needs of the second-phase survey. Some of this is driven by disproportionate allocation in the second phase, which may use up a large proportion of a particular subgroup, which can also preclude the first-phase sample's use by other surveys. Small-area estimation is another hurdle for second-phase samples.

All of these factors lead to a set of administrative challenges that have been briefly mentioned in the context of the examples provided but are worth acknowledging more generally, Lepkowski said. One such challenge involves funding, particularly deciding on how the second-phase survey can share some of the costs of the first-phase survey (e.g., the costs related to screening or list-ing). Another challenge is related to the changes in sample size and the logistics associated with adapting to these changes. Second-phase surveys tend to be administered after the first survey, although concurrent designs are also pos-sible, and these represent a separate set of administrative challenges. The use of some sample frames, such as the Master Address File (MAF), has limitations that impose restrictions on second-phase survey operations.

Something that is not typical of currently existing two-phase surveys is a conscious effort to design them as true two-phase surveys from the outset.

Instead, second-phase surveys seem to occur on the basis of arising opportunities for collaboration between agencies and an after-the-fact recognition that there is a possibility to save on costs across two or more organizations.

A joint design from the outset would allow for optimal allocation across phases and better input into units of selection. Two-phase surveys could also be more successful at reducing nonresponse if the goals and designs of both surveys were kept in mind. This would allow for the planning of a more comprehensive incentive structure, as well as tracking and follow-up procedures. There is also tremendous opportunity to use paradata and a responsive design for utilizing first-phase data to predict what will happen in the second phase. Prediction models compared with what actually transpired in the second phase can then be used, improving the ability to intervene and improve response properties.

THE POTENTIAL ROLE OF THE AMERICAN COMMUNITY
SURVEY IN SAMPLING RARE POPULATIONS

Keith Rust (Westat) began by saying that he added the word "potential" to the title of his presentation to illustrate that some of the ideas presented are in development or are under consideration, rather than already in progress. He then proceeded with an overview of the American Community Survey (ACS).

Conducted by the Census Bureau, the ACS surveys approximately 250,000 households each month by mail, or 3 million households per year. The questionnaire contains 48 questions about each individual in the household and 21 questions on housing. Nonrespondents to the mailed questionnaire receive a telephone follow-up whenever possible (when a phone number is available). The remaining nonrespondents for whom there is no phone number or who did not respond by phone are eligible to be in the sample for follow-up by an in-person interview using computer-assisted personal interviewing (CAPI) technology. The in-person follow-up obtains interviews from about one-third of the 48 percent of nonrespondents who do not respond by mail or telephone. But the CAPI subsample rate does vary by population group.

The overall weighted response rate to the ACS is very high at 97-98 percent, but due to CAPI subsampling for follow-up, the data actually obtained are about two-thirds of the original sample. Therefore, data are obtained for approximately 2 million households per year. Differential sampling also affects the total final count of respondents. The sampling for the ACS is complex, but, as an example, there is an initial oversample of small governmental units. This works out to about 15 percent of the sample, which covers 5 percent of the population in these units. Also, since nonresponse CAPI subsampling yields about one-quarter of the sample that is obtained through CAPI, these interviews get three times the weight of the remainder. This suggests that the effective sample size due to the differential weighting is closer to 1.5 million

household interviews per year, although the design effects due to weighting could vary among subgroups.

As with any survey collected by mail, there is item nonresponse. There are a lot of questions on the ACS, and some of them are open-ended responses that must be coded (e.g., industry, occupation, field of degree). There is also the issue of response error, particularly when it comes to reporting income. Some questions involve a challenging recall task, such as the question about employment. Each of these factors can contribute to item nonresponse and response error.

It is in this context that the use of the ACS as a frame for sampling rare populations should be considered, Rust said. Issues to keep in mind with sampling rare populations are cost and burden of sampling, timeliness of the data available, the sample size available, the amount of cumulation that is needed (from the ACS), the effects of differential weighting, coverage issues, response error, the quality of the contact information, sampling error estimation, and confidentiality and human subjects concerns.

One of the most obvious benefits to using the ACS as a frame for other surveys is the reduction in the cost and burden associated with smaller surveys. Cost is reduced for the smaller survey by not having to screen a large initial sample in order to identify a subpopulation of interest. Respondent burden is reduced by not having to participate in a screening survey. Furthermore, there is the ability to fine-tune sample allocation for different population subgroups. Sample size can also be controlled precisely because the sampling done is from a frame of people known to be in the population of interest. Finally, it is possible to orchestrate the release of sample in waves or replicates in order to fine-tune yield.

As Lepkowski mentioned in the previous presentation, the timeliness of data available for use as another survey's sampling frame is also a consideration, Rust said. In this case, what proportion of people will have a status change that might cause them to move into or out of the population of interest? As an extreme example, the ACS would be of no use as a frame in the case of newborns, very recent immigrants, or the recently unemployed. Another question is what constitutes a sufficient sample for the rarest group of interest. If cumulation of data over many months or years is required, then issues of timeliness are exacerbated. Furthermore, the differential representation in the ACS sample may lead to large weighting design effects in a rare population, although some of this may be offset with subsampling—if there is enough sample to do this.

Like most surveys, the ACS probably undercovers certain groups (potentially the groups of interest) in the population. Data from the census undercover newborns; it is likely that the ACS does as well. Household surveys tend to undercover young adult black men, so it seems likely that the ACS would, too. The ACS weighting adjustments can help address undercoverage for estimates, but it is unknown how useful this will be for the subsampled rare population group.

Misclassification as it relates to rare population status can result in substantial undercoverage and wasted sample, Rust went on. Any survey of a rare population that uses screener identification will have this problem. Furthermore, in the case of the ACS, which is largely a mail survey, there is no interviewer who can follow up with probes to ascertain that a respondent is answering a particular question correctly.

The quality of the contact information that is available on the ACS is another issue to consider, Rust observed. Is the address information on the ACS accurate enough for follow-up by mail, telephone, or in-person contact? The ACS does not ask for address corrections or clarifications on its form. This could be a potentially significant issue, particularly for multiunit structures, he said. If the contact information is sufficient for a subsample, there is the related issue of confidentiality and human subjects protection issues. The ACS response is required by law; respondents are told that their responses are confidential and will be used for statistical purposes only. Title 13 of the U.S. Code, which authorizes collection of personally identifiable information, requires that follow-up surveys must be conducted by the Census Bureau because the information collected in the ACS is confidential. Thus, access to this information cannot be shared outside the agency.

The ACS sample is a rolling sample, with a new sample produced every month. Could this be utilized to design rolling samples for rare populations? It may be possible to draw sample from the ACS every quarter, but, for reporting subgroups, data can be cumulated across quarters to get a continuous rolling sample. This could be used to measure trends, Rust said.

Another question that arises is whether the ACS in its own right is sufficient to identify a rare population of interest. This suggests the possibility of adding questions to the ACS to be used as a screener for identifying a rare population. This leads further to what kind and how many questions can be asked, as well as who is responsible for the quality of the data from these questions. He said it is important to distinguish screener questions from those that will be tabulated along with other ACS data. How will the effect of adding questions to the ACS on response rates be evaluated? He observed that this may not be the right time to add questions, given suggestions that the ACS should be cancelled altogether, or at least made voluntary, because of claims that the survey is too intrusive.

Rust noted that a couple of examples can be used as case studies of smaller surveys using the ACS for sample creation. One is the National Science Foundation's National Survey of College Graduates (NSCG). This survey, conducted by the Census Bureau in the past, measures the number and characteristics of people with science and engineering degrees. Formerly the frame for the NSCG was the census long-form sample. Since the long-form sample no longer exists, the ACS will be used as a frame instead. A "field-of-degree" question was added to the ACS specifically for that purpose

(although it is also of interest in its own right). The benefit of adding the question is that it permits oversampling of people with science and engineering degrees. However, several years of ACS data will be required to achieve what has previously been the desired sample for some of the groups. Still, this is a vital question for targeting the sample of persons with science and engineering degrees, and getting that information from the ACS greatly decreases screening costs. The field-of-degree question does have its problems, he said; it is an open-ended question and therefore requires extensive coding. And in 2009 there was 9 percent item nonresponse. There are most likely issues of data quality and coverage. And this also raises the question of whether the NSCG could benefit from using a rolling sample, at least for a component.

The second case study describes a test of the feasibility of using the ACS for the National Immunization Survey (NIS). The NIS produces annual vaccination rates for children ages 19 to 35 months, plus a component for teenagers ages 13-17 years. It produces data at multiple levels of geography, including 78 areas known as Immunization Action Plan Areas. The NIS currently uses a list-assisted random digit dialing (RDD) sample—a methodology with high screening costs, because only 5 percent of households have infants. And the sample size is quite large: 26,000 infants per year and 31,000 teens.

Rust observed that this survey, like others, experiences many of the problems associated with telephone surveys: low response rates and undercoverage, to name just two. To help combat these problems, the proposal was to consider using the ACS as a frame for the NIS. The ACS certainly offers the possibility to overcome many of the current deficiencies in the NIS sample, and the idea of a rolling sample would integrate naturally into the NIS design. There are also rich data on respondents that could be used for adjustment and bias analyses. The ACS probably undercovers persons under 1 year of age, so there are probably coverage problems. The immunization surveys are interested in children ages 19 months and older, but because of the time lag, those under 1 year of age would need to be selected from the ACS. Moreover, the NIS would need to be in the field within 19 months of the ACS response to cover 19-month-olds.

The Census Bureau and the Centers for Disease Control and Prevention jointly conducted a one-state trial with children ages 19-35 months using ACS data for the period 2006-2008. They found that although the response rate was good, in-person interviewing was vital. A provider check was included in the survey, in which respondents gave contact information for those who provided the immunization. Generally, respondents gave good information about the provider, but confidentiality issues were raised related to the fact that the respondents were identified on the basis of the ACS. As a work-around, Rust said, providers were given special sworn status by the Census Bureau. Although this appeared to work for the trial, it may be an issue for surveys that want to use the ACS as a frame.

The ACS has the potential to greatly reduce screening costs and reduce

undercoverage for surveys of special populations. It is also rich in respondent information, which is quite useful for enhancing estimation. These benefits may be overstated, however, as the evidence suggests that a significant amount of in-person interviewing will be needed. Other issues include timing, coverage, data quality, sample size, and differential weighting. Using the ACS data as the basis for a person's inclusion in a future survey raises issues of consent and confidentiality. Issues related to including additional questions on the ACS (such as how many additional questions and who decides on the questions) are other hurdles that any survey using the ACS as a frame must overcome. Given current Title 13 restrictions, the Census Bureau must conduct the survey and maintain the data. Rust ended by saying that, although the ACS appears to offer opportunities for use as a sampling frame for other surveys, it is not a panacea, and there are real risks of abuse.

SAMPLING FRAMES FOR FEDERAL HOUSEHOLD SURVEYS: A VISION FOR THE FUTURE

Frederick Scheuren (National Opinion Research Center) began his talk on sampling frames by saying that the goal of the workshop should be to identify ways of supporting an information society, not just individual agency missions. The focus should be on multimode and multiagency sampling frames. He noted that even the concept of "household" survey frames is too narrow and unable to describe many new developments, such as the spread of cell phones.

Scheuren said that government agencies do not typically cooperate well, except in times of crises. But there are some common challenges across the federal statistical system that need to be addressed: survey costs are too high and the delivery of information is too slow. Referring back to the presentation by Rust, he gave the example of outdated sampling frames that can be so old by the time they are shared with another agency that they are no longer useful.

Scheuren argued that government data collections will become increasingly difficult to conduct in a data-rich world, with information becoming available from many competing data sources. This means that government agencies will have to learn to adapt or they will risk irrelevance. Some of the examples from other countries discussed earlier are cheaper, faster, and more responsive, and they should not be ignored, he said, even if the same approaches cannot be implemented in the United States.

A possible new paradigm in a data-rich world could be characterized by emphasis on paradata, both design and model-based estimation, and quick, simulated outcomes, instead of traditional estimators. Statistical systems can no longer afford large samples, so small sample properties also have to be stressed in this context.

Unified sampling frames are an important consideration for the future. These could be assembled starting with geographic addresses, which would then be

linked to sampling frames based on other modes of contact, such as telephone (both landline and cell phone) and web. An ideal unified sampling frame would also integrate information available from such sources as the census, the American Community Survey, the Internal Revenue Service, the Social Security Administration, and the U.S. Postal Service. The key benefit to adding the administrative data is that they enable stratification of the sample on variables of interest.

Scheuren argued that cooperation related to work on sampling frames is important, and that in his view all federal survey contracts should require data sharing for frame construction. He acknowledged that there are many barriers to implementing this type of mandatory data sharing, including legislative and regulatory restrictions, the lack of political and bureaucratic will, the fact that it violates precedents, and the need for a long-term commitment. Confidentiality concerns could be addressed if access is through a data enclave—in other words, a secure environment that provides authorized researchers access to confidential microdata. Access would have to be limited in purpose to frame construction only and subject to oversight by a neutral entity, such as the Office of Management and Budget.

Given the large opportunity costs, it is important to consider whether implementing a unified sampling frame is worth it, said Scheuren. Such a system could improve data quality and enable faster delivery times. However, maintaining the frame will be expensive, and depositing the data in the enclave will take time, which means that at least some of the information will be outdated. He noted that although an obvious benefit of the work on a unified sampling frame would be the development of a cooperative structure in the federal statistical system, a large investment in sampling frames could turn out not to be the right long-term investment.

Scheuren reiterated that the world is becoming more data dense, and government statistical agencies now have strong competitors. This could mean that, in the future, surveys and censuses will have a smaller role, and emphasis will shift to increasing reliance on administrative data and to combining information, which is more than just combining data. The federal statistical system will be ready for this new reality if agencies invest in becoming more cooperative.

DISCUSSION

Frauke Kreuter (University of Maryland) asked whether Rust had a sense of what the cost savings would be for the NIS if it were to use the ACS as a sampling frame. Rust responded that although he did not know, he did not think cost savings were a particular consideration; this initiative was probably driven by quality concerns and dissatisfaction with how the survey is conducted now. Marcie Cynamon (National Center for Health Statistics) added that there were concerns about the NIS coverage related to the percentage of the children who were in the age range through the RDD and cell phone components.

Related to the ACS as a frame for the NSCG, Robert Kominski suggested that the argument for including a screener question on the ACS for the NSCG was the cost savings of $3 to $6 million. There was a good chance that, in many cases, the addresses in the NSCG frame were no longer applicable for persons who held at least a bachelor's degree. However, even with the delay in ACS processing, the data would still be more current than alternatives. In addition, getting data more frequently makes the sample to draw from much larger, even though the sampling rate is much smaller.

Cynamon noted that the National Health Interview Survey gives half of its sample to the MEPS, and it is not a cherry-picked half. Trena Ezzati-Rice (Agency for Healthcare Research and Quality) discussed the screening efficiencies that are gained from the integration of the NHIS and the MEPS. The integration has been extremely helpful in the benchmarking of the survey estimates, she said, and this has been useful for both surveys. The end result of the NHIS-MEPS integration is a very rich frame of auxiliary variables that can be used for nonresponse adjustments. Further research has found that incorporation of some of the health variables from the NHIS reduces bias in the MEPS estimates.

Lepkowski added that, if there is this rich set of data and methods for doing model-based estimates on something like the MEPS and the extension of the frame to the other half of the NHIS sample, then why are the only estimates produced based on the MEPS sample? The estimates could be of higher precision, but currently these resources are not being utilized, which is part of the failure of two-phase designs.

Sondik added that, beyond the use of a survey as a sampling frame for another survey, the potential is also there for using substantive data from the first survey, although implementing this link would require substantial resources. Ezzati-Rice commented that the integration of the NHIS and the MEPS has provided an additional data point—beyond the two years of the MEPS data—for longitudinal analysis. There has not been as much mining of these data as there perhaps could be, but they have been looked at, specifically the transitions in health insurance coverage from the NHIS to the two panels of the MEPS. There is also an interagency agreement to look at cancer survivors from the NHIS supplement and then to look at issues related to health care access and costs for these respondents in the MEPS.

On recycling sample, Kominski said that currently an ACS case can be used in only one secondary survey based on the primary survey, but this is not necessarily the case in the commercial sector, and other scenarios could be considered. A participant from the National Agricultural Statistics Service added that they have this problem with the Agricultural Resource Management Survey: first, there is a screening survey, and then another survey uses up the data, so they cannot be used for the next survey. A used case is taken from the frame, and the weight of everything else is increased proportionally. The ATUS adjusts its weights similarly.

Another speaker commented on the importance of developing a process of selection for who gets to use what portion of a sample. What are the terms of the agreement? And how does this get resolved?

Citro said that the ACS has the benefit of containing questions that can be used for screening and weighting, but the available sample is limited. The MAF does not contain survey responses, but the efficiencies associated with sharing it would be a major step forward for the statistical system. She added that although the sharing of the MAF is limited by Title 13 of the U.S. Code, it is conceivable to think of enabling the sharing of some version of the MAF. For example, one of the concerns is that, as part of its work, the Census Bureau discovers sensitive information, such as a building that contains more housing units than it is supposed to. However, this type of data could be collapsed across cases. Under the Confidential Information Protection and Statistical Efficiency Act (CIPSEA), a "slightly sanitized" version of the MAF could be entrusted to the statistical system as a whole, and federal agencies could collaborate, perhaps in the context of a data enclave, on building, validating, and using the MAF for sampling frames. As part of the data enclave, access could potentially be also extended to data collection contractors who work on federal surveys.

Kalton commented that the opportunity costs associated with Scheuren's vision are high, and the approach described by Citro is more manageable. It is not only the statistical system that could benefit from access to the MAF, but the MAF would also benefit from more agencies working on improving its quality.

Sondik asked Scheuren about his vision for the federal statistical system and the idea of increasingly fuzzier data and methods: How will it be decided what constitutes a benchmark? Scheuren responded that it is critical to take advantage of opportunities that are already available: administrative records, business frames, modern methods, sharing mechanisms, and the knowledge of those who move around and have worked in other countries—especially smaller ones forced to use more economical means to obtain data. Finally, he said, the cooperation and common culture that existed in the Census Bureau and unified the system in the 1940s and 1950s could be resurrected.

4

Collection of Household Data

NEW DATA COLLECTION MODES AND THE
CHALLENGE OF MAKING THEM EFFECTIVE

Don Dillman (Washington State University) began his presentation by saying that surveys are now in a period of tailored design, in which different modes and implementation procedures are appropriate for different situations. The focus of his talk was on the challenges this new period presents.

An understanding of the evolution of survey modes and designs is important for gaining perspective on the current system. First was the transition from face-to-face interviews to telephone. Dillman recalled that his own experience with this transition was relatively easy, because face-to-face methods could readily be adapted to telephone surveys. Households had about 78 percent telephone coverage in 1970, and this number seemed to be increasing, making the transition increasingly more feasible.

Household survey methods, including sampling approaches, could reasonably be applied to telephone, he said. The interviewer's role in telephone surveys is similar to that in a face-to-face interview in terms of reading items, clarifying questions, and relying on hidden categories (categories that are not offered to the respondent), as needed. The main differences are that show cards need to be eliminated, scales have to be shortened to achieve the same level of comprehension, and questions sometimes need fewer words to be understood aurally. Another difference is that supervisors are more accessible during telephone than face-to-face interviews.

Bringing email and the web into data collection was a more difficult transition. Currently, approximately two-thirds of households have Internet access

35

and use it with some regularity, leaving a possible one-third of households unable to respond to a survey over the Internet from home.

Another problem that arises with creating a sample of Internet respondents is that it is harder to implement a within-household random selection because some householders lack Internet skills. In the case of some households, this phenomenon may be related to a division of labor: just as some people do the laundry and some take care of cars, a particular person in a household may use the Internet. Furthermore, survey organizations generally do not have email addresses that would enable them to send respondents links to Internet surveys, unless a prior relationship exists. Even if this could be resolved, it is likely that response to an initial email invitation would be quite low.

Meanwhile, the telephone is losing its viability as a survey mode option. There are many reasons for this, including the increasing use of cell phones (although these can sometimes be added to a frame), the decreasing reliance on landlines (current coverage is less than 75 percent of households), and increasingly blurred lines when it comes to the geography of phone numbers. American culture has also changed. People no longer use the telephone for most business interactions unless they have to, and they tend to exercise more control over their devices than in the past, by not always answering calls.

The telephone itself now fulfills a variety of functions, often serving as a personal computer. However, the screen space available for a web questionnaire is small, and entering text on a telephone is prone to error. Finally, responding to a survey on a phone device often cannot be combined well with other activities the potential respondent may be doing while accessing the Internet.

Changes related to the telephone and the continuing limitations of Internet access suggest that, in the near future, there will be more reliance on mixed-mode survey designs to collect data. Dillman devised a typology of the ways data collection modes are most commonly mixed (Dillman et al., 2009), summarized in Box 4-1.

The first type involves the use of a particular mode to encourage people to respond by another mode (typically, the Internet). In a sense, this is still a single-mode study, and therefore measurement differences between modes are not as big a concern as they might be otherwise. In the second type, one mode is used to ask some of the questions, and another mode to ask others, such as more sensitive questions. In practice, this interview technique often entails an interviewer simply turning a laptop around during a face-to-face interview so that the respondent can self-administer part of the interview. A third type of mixed-mode design involves using different modes of administration for different types of respondents. A fourth approach, typically used in longitudinal studies, employs one interview mode for the first interview and another mode for the second and subsequent interviews.

Dillman pointed out that it is important to remember when combining different modes of administration that sometimes achieving one survey objective

---

BOX 4-1
Typology of Mixed-Mode Surveys

Type 1: One mode for data collection, another mode for selection/encouragement.

Type 2: One mode to ask certain questions, another mode for additional questions.

Type 3: One mode for some respondents, another mode for other respondents.

Type 4: One mode for Time 1 data collection, another for Time 2 data collection.

SOURCE: Workshop presentation by Don Dillman.

---

may get in the way of another. For example, improving response rates by offering alternative modes of responding may introduce measurement differences, or reducing costs may conflict with obtaining quicker responses.

There are also several significant barriers to wider adoption of mixed mode designs, he said. There is a tendency among survey professionals to construct survey questions differently for different modes, and part of the reason for this is the desire to maximize the design for a specific mode. Visual (self-administered) versus aural (telephone) presentations, in particular, have different requirements.

For example, in the face-to-face mode, show cards can be used for answer choices, scales are often fully labeled, questions and questionnaires tend to be longer, and some of the answer options can be made available to the interviewer without explicitly offering them to the respondent (such as "Don't know" or "Refused"). In the telephone mode, scales tend to be shorter and are presented without all categories labeled, questionnaires are shorter, complex branching formats can be used without affecting respondent comprehension, and, as in the face-to-face mode, some answer options can be made available without being explicitly offered. The mail mode encourages less question branching but can accommodate longer, more complex scales. Open-ended question formats are avoided when possible, and response categories cannot be hidden. A web mode encourages required answers and fewer "don't know" options. Fill-ins are possible from previous answers. Audio, video, and other add-ons are possible, and typically there are no hidden categories. Unintentional mode-related construction differences can often lead to significant differences in the distribution of the answers provided.

Research has shown that the visual layout of survey items influences answers. Dillman highlighted the 24 most significant concepts in visual design

(see Box 4-2). As an example of different design requirements for visual communication, he described a challenge encountered by the National Science Foundation while designing one of its web surveys. The goal was to obtain date information from respondents using two digits for month and four digits for year, in adjacent character spaces. Cognitive interviewing revealed that respondents will attempt a variety of approaches to answering a date question (e.g., using alphabetic abbreviations for the month) and that they get frustrated when

---

BOX 4-2
Visual Design Concepts That Matter

Attention and visual processing:
  Preattentive processing
  Attentive processing
  Useful field of view
  Foveal view
  Top-down processing
  Bottom-up processing

Visual features that influence the expression of words, numbers, and symbols:
  Figure/ground composition
  Size
  Shape
  Location
  Spatial arrangement
  Color
  Brightness
  Contrast

Languages that give independent meaning to information on a page:
  Words
  Numbers
  Symbols

Grouping principles:
  Pragnanz (law of simplicity)
  Proximity
  Elemental connectedness
  Common region
  Continuity
  Closure
  Common fate

SOURCE: Workshop presentation by Don Dillman.
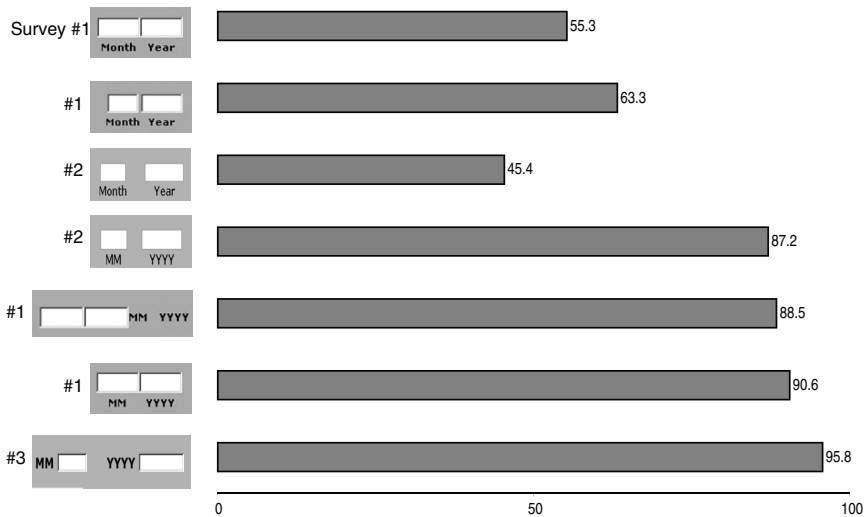
FIGURE 4-1 Summary of web experiments.
SOURCE: Workshop presentation by Don Dillman.

they receive an error message. This led to extensive testing of this question over a period of four years.

Figure 4-1 shows that changes in visual formatting led to large differences. According to the law of proximity in Gestalt psychology, if something is connected, it tells people to do the same. When this principle was applied in experiments, 55 percent of respondents filled in the boxes correctly. If the month box was smaller and the year box a little larger, 63 percent filled in the boxes correctly. When the symbolic language MM, YYYY was added to the respective boxes, this yielded 87 percent correct responses. Finally, when boxes and symbolic language were arranged in natural reading order, 96 percent of respondents provided responses in the desired format.

Dillman also described some experiments to address the issue of visual versus aural presentation. In one study, he asked respondents in three different ways when they began their studies: (1) When did you begin your studies? (2) What date did you begin your studies? And (3) What month and year did you begin your studies? On the web survey, there was little difference in the percentage of students using the preferred MM/YYYY format. However, over the phone, the differences between the distribution of the responses were drastic. The percentage of respondents reporting month and year was 13.4 in the "when" condition, 49.5 in the "what date" condition, and 83.7 percent in the "what month and year" condition.

Of course, in the case of telephone interviews, the interviewer can act as an "intelligent system" that converts the responses to the desired format. That luxury does not exist in a web mode, forcing researchers to think of questionnaire construction differently and to invoke theoretical concepts on visual information processing.

Another issue related to different modes of administration involves scalar questions. The concepts of social desirability, acquiescence, primacy, and recency have been often used to explain why people respond the way they do, but Dillman argued that these concepts often do not explain mode differences. He and colleagues conducted several experiments to examine whether using the same wording for scalar questions will produce the same answers in aural as in visual presentations. The experiments involved a variety of scales, including 5-point, 7-point, fully labeled, and polar point labeled scales. Regardless of the scale type, each of the experiments resulted in slightly more positive responses on the telephone than on the web. The point here is that there is a consistent body of evidence building that mode makes a difference in responses.

A line of research Dillman is particularly interested in involves combining two visual modes of data collection and avoiding the aural mode. Sending an email request as a first contact is typically not appropriate in cross-sectional household surveys, unless there is an existing relationship with the sample members or if they are a part of a longitudinal study. When given a choice of mail or web response, through mail contact, people tend to opt for mail, and overall response rates are lower. Requests for web-only responses typically result in low response rates.

However, and despite declining response rates for most modes of data collection, response rates in mail surveys, particularly with prior screenings or incentives, tend to remain fairly high. Some of the reasons can be explained by social exchange theory, and such concepts as rewards/benefits, burden/costs, and trust in the delivery of benefits. Social exchange theory could serve as a guide for other self-administered modes, such as the web, and for mixing modes in order to avoid having to rely on email only to obtain web responses and postal contacts only to get postal responses.

In many ways the Internet is different. There are problems with using it for surveys: the burden can be greater when responding to a survey via the Internet, particularly if going from postal letter to the computer; computer literacy is low for some respondents; there are operational issues—Does the computer work properly, or at all?—and emails from strangers can be harder to find or get lost more easily after the first day in one's email inbox.

The benefits of Internet surveys vary. Technology is easier to deal with for some than for others. For some, there may be faster ways of responding. With an Internet survey, there is no need to try to find a mailbox to return a questionnaire. But with Internet surveys, trust is a significant concern. People do not like to open email from strangers, the sources of emails and websites can be faked,

and there is the ever-present threat of downloading a virus or other malicious software. This last issue represents an area in which government agencies may have an advantage: people tend to trust communication coming from a government authority much more than any other potential survey contact.

Still, if people are given a choice of responding by either mail or Internet, most will chose mail. And, if mail is withheld to encourage respondents to use the web, research has shown that the respondents who end up participating during follow-up are very different from one mode to another. Dillman noticed in his research, however, that if an address-based sample is used to try to push people to the web, the result is a greater response from an advance postal token incentive for the mail-plus-web combination than for just the mail response alone. Email tends to cut the burden of web response because it brings respondents closer to their response mode preference. In essence, what will best bring postal, email, and web contacts together to obtain more responses by web is to begin integrating two modes, rather than forcing all web options together or mail options together.

In Dillman's view, it is important for the survey community to bring together token cash incentives, mode choice, and email augmentation in trying to move forward. New options like address-based sampling and the sequential use of modes need further exploration but have great potential.

He ended by saying that the transition to the web is desirable, but it is going to be difficult. A positive development is that Dillman's experiments that were based on address-based samples have yielded two-thirds of the responses over the web, which three or four years ago would not have been possible. However, coverage limitations suggest the need to use another mode (most likely mail) to at least deliver the request. This also raises concerns about mode differences. Evidence is mounting that the aural and visual modes sometimes produce different responses.

INTEGRATING ADMINISTRATIVE RECORDS INTO THE FEDERAL STATISTICAL SYSTEM 2.0

The focus of the presentation by Rochelle Martinez (Office of Management and Budget) was to illustrate what the statistical system could do to address barriers to making greater use of administrative records. For the past few years, interesting work has been going on to try to build capacity to use more administrative records, particularly with demographic data collection. Her talk specifically addressed the work going on across the statistical system, coordinated by the Office of Management and Budget (OMB). She discussed initiatives in the president's budget and recent events related to administration support for these activities.

For many years, members of the statistical community have said that administrative records can and should be used more fully in the federal statis-

tical system and in federal programs. The use of administrative records in the Netherlands and other countries gives a good flavor of the kinds of things the statistical system can envision doing in the United States to varying degrees. There are also areas, however, in which substantial work has already been done in the U.S. context. Most notably, administrative records have been used in economic statistical programs since the 1940s. There are also good examples of administrative data use with vital statistics, population estimates, and other programs across several federal statistical agencies.

Martinez mentioned that former director of the U.S. Census Bureau, Kenneth Prewitt, often talks about another reason that administrative records hold potential: the need for innovation. He has said that he is less concerned about the federal statistical system with regard to relevance and integrity than he is about innovation, in particular about how prepared statistical agencies are for the innovation necessary to navigate the new world. In many cases, national information systems are increasingly reliant on administrative data and, in some instances, on data from the commercial sector. Prewitt's greatest concern is that government agencies seeking statistical information about the population will bypass statistical agencies altogether as they turn to the parts of the government that control large administrative data sets.

Martinez said that she sees this happening in some federal agencies right now. Offices that are collecting data for administrative purposes can (at least reportedly) produce a statistical result much more quickly than the principal statistical agency in that department. For a congressional or public affairs office, this is very appealing. Those in the statistical system can think of reasons why that might be a problem, but these offices may not. The best case scenario is that there are multiple estimates in the public domain that somebody has to be able to explain. The worst case is that somebody thinks that a statistical agency is less relevant and less timely and therefore that its data are less useful than the administrative data source. At OMB, Open Government and Data.gov initiatives encourage putting many more administrative data sets in the public domain, where they can be used for a variety of purposes, so these issues need to be addressed across the system.

Members of the Federal Committee on Statistical Methodology (FCSM) wanted to facilitate statistical agency use of administrative records. To explore how to achieve this, an interagency subcommittee was formed. This group created a set of products that the statistical community may find useful going forward.

The first product to come out of the subcommittee was a set of case studies, "Profiles in Success," focusing on projects that had successfully acquired and used administrative data in a statistical project. Martinez said that the case studies were quite useful in helping the subcommittee members identify systematic barriers to greater use of administrative records. It is these barriers that the group has tried to address head-on in recent months and years.

Following the "Profiles" product, the subcommittee turned to awareness activities, in part to dispel myths related to difficulties related to using state administrative records data. This group found many good examples of successful administrative data use in research and, in some cases, production. The subcommittee wanted to highlight the necessary success factors for using administrative data, and the statistical community has been very receptive. As a result, the subcommittee has been asked to develop training and other activities to help data users navigate the difficult world of acquiring and using administrative data.

A subsequent product for the toolkit, she said, was one of creating model agreements. Getting an agreement in place for data sharing and usage between agencies is often a drain on time and money. Thus, the subcommittee has created a model agreement that agencies can use to facilitate the data-sharing process. Although many aspects of such agreements can be covered in a template, not all can, so there will be tailoring to some extent. The idea behind model agreements is to reduce front-end costs, because so many projects either die on the vine at this stage or use too many project resources, leaving fewer resources for the research.

Another product created by the subcommittee is related to informed consent. The informed consent product is an in-depth look at legal requirements across federal agencies, current practices for informed consent at statistical agencies, and current practices at administrative agencies. It also synthesizes research on informed consent wording in the context of data sharing and record linkage. This product is likely to help the statistical system in terms of best practices for new activities going forward. It will also provide guidance on how to meet requirements for projects for which administrative data were collected before there was an identified statistical use for them. The subcommittee has also done some work on data quality, with the goal of creating tools for data quality measurement and documentation, but it is far from complete.

As a result of the subcommittee's work, Martinez went on, at least four barriers to using administrative data crystallized. One of these barriers is statistical agency access to administrative data. Statistical agencies have statutes that are designed to protect the confidentiality of data, and they consider themselves very much stewards of data. But despite these provisions and helpful language in the Privacy Act, statistical uses of administrative data are sometimes difficult to achieve. In many departments, program offices have data on which the legislation is either silent, unclear, or perhaps narrow in terms of the kinds of uses that are considered appropriate.

There is also an issue of incentives; program offices may not think it worth the effort to figure out how to address a statistical agency's request for data. Whose job is it to work with the statistical agency? It can be very time-consuming to identify variables that are needed or to work with an agency to understand what data they have now or how these could be used. Negotiating

agreements is a practical product that comes out of these discussions. Some agencies spend years and years trying to obtain administrative data. Statistical agency access to administrative data may be the most important barrier because, without access, projects cannot be undertaken.

A second, somewhat related barrier is what the subcommittee has termed inadequate infrastructure, referring to the infrastructure at both the statistical office and the administrative office. There is an administrative infrastructure needed to address such issues as the process for requesting data and approving the request. Technical infrastructure can require a significant investment of time and resources on the statistical side. But even on the administrative agency side, someone has to be able to extract and transfer the data. The subcommittee thinks that infrastructure is lacking in many of these cases.

The third barrier is administrative data quality. Although they are not perfect, with survey data, agencies have the capability to describe and to understand the quality of what they have. In other words, there are a lot of measurement tools for survey data that do not yet exist for administrative records. Some have assumed that administrative data are a gold standard of data, that they are the truth. However, others in the statistical community think quite the opposite: that survey data are more likely to be of better quality. Without a common vocabulary and a common set of measurements between the two types of data, the conversation about data quality becomes subjective.

Another significant data quality issue for statistical agencies is the bias that comes with the refusal or the inability to successfully link records. In addition to the quality of the administrative data as an input, the quality of the data as they come out of a linkage must be considered as well.

The final barrier has to do with researcher access. This includes researchers both internal and external to the government. Sometimes an afterthought, this is the idea of creating documentation that would be needed to really make a file, particularly a linked file, useful for someone else outside the project. There are issues of documentation and of providing disclosure protection to a linked file. For this reason, linked files are very rarely public-use files. Few methods for restricted access have been devised beyond those that existed for projects before record linkage was a focus. Many of these linked files have been created and not really used by people outside the immediate project, and that is a concern both in terms of the utility of what has been created and for data quality.

Martinez said that some initiatives in the president's fiscal year 2011 budget should help further the subcommittee's goal of promoting the use and exchange of administrative data. Specifically, three major pilot studies have been proposed, two for the Census Bureau (2010 Census Simulation Pilot and Health Data Pilot) and one for the Economic Research Service (Nutrition and Food Assistance Pilot).

Together, these three pilot studies are designed to address all four barriers. Although the barriers will not be resolved in a year, agencies can certainly

begin to address them in ways that benefit the entire federal statistical system. Martinez emphasized that the notion of a common good was very important in proposing the initiative.

The first pilot project is designed to use both government and commercial administrative data to see if it is possible to simulate 2010 census results. Outcomes envisioned include advancing both knowledge about and measurement of the quality of many administrative record data sets. Ideally, this will not only inform the decennial census, but also other demographic surveys.

In Martinez's view, this project is also critical to setting up an infrastructure. Some consider the Census Bureau to be the ideal place for this, because it is thought to be big enough and stable enough to handle a large number of different files and many different activities. This is why the Census Bureau also received much of the funding; it would be much less efficient to attempt to build up infrastructure at multiple statistical agencies than to centralize the technology, capacity, expertise, and synergy.

The second pilot project is related to the first one and is also housed mostly at the Census Bureau. The idea is that the Census Bureau has the capacity and stabilizing infrastructure that enables it to provide record linkage services to other federal statistical agencies. The National Center for Health Statistics (NCHS) has agreed to be the pilot agency to provide identifiers from multiple health-related administrative and survey data sets to the Census Bureau to link and return to NCHS.

The overarching concept behind this pilot study is that record linkage is a service, a line of business that the Census Bureau could provide to agencies that are smaller or that lack similar capacity. A vision for the future is to centralize to some degree the expertise and the hands-on experience with different data files while still retaining the benefit of having a subject-matter agency, such as NCHS, getting back the data and using them for both subject-matter research and for providing access to other health researchers.

The goal of the third pilot project, the nutrition project, is to help the statistical community better understand how to acquire and use state administrative records for statistical research and to demonstrate the utility of such data for program evaluation. The hope is that this project can help identify a model in which these data might be acquired in a more centralized way. This project also helps to bring together multiple agencies that are interested in state data.

Although a primary goal of the pilots is to address the barriers outlined, Martinez said that these projects have also created interest among policy officials because of the ability to learn more from a subject-matter perspective. To make any of these ideas happen, it is essential that administrative agencies be included in the conversations about these uses of their data.

To that end, OMB has recently issued a memorandum encouraging federal agencies to share data in order to meet the needs of several administration initiatives, including statistical data projects. This demonstrates that administra-

tion officials are supportive of these efforts to increase the use of administrative data. The support of senior officials will be necessary, she said, because a move to expand administrative data use necessarily entails difficult conversations about legal and policy issues regarding data access.

Martinez added that all of the work she described was sponsored by the Interagency Council on Statistical Policy (ICSP). The ICSP is comprised of the heads of the principal statistical agencies. Among these agency heads a subgroup has been focused on developing a vision beyond the three pilot projects. She said that among agency heads and project teams alike, there is continued enthusiasm for these projects, and they are hopeful that the studies can continue to move forward in an uncertain budget environment.

Despite operating under a continuing resolution, project teams have already been working on the aforementioned pilot projects. These groups would like to involve more researchers in the projects to help think through some of the issues that crop up in the course of the work. Furthermore, it is very important that not only federal statistical agencies, but also the professional statistical community, and particularly those working in the states, contribute to this conversation.

THE ROLE OF ADMINSTRATIVE RECORDS IN HOUSEHOLD SURVEYS: THE CANADIAN PERSPECTIVE

Julie Trépanier (Statistics Canada) described her agency's use of administrative records in household surveys. To set the stage for this perspective, she outlined official legislation, policies, and guidelines that govern administrative data use in Canada.

Statistics Canada's guiding principle—though not a policy—is to use administrative records whenever they present a cost-effective alternative to direct data collection. Section 13 of the Statistics Act allows Statistics Canada to obtain administrative data files from any organization for the purposes of the law. It also specifies some rights of access to administrative data. Specifically, Section 24 gives Statistics Canada the right to use income tax records; Section 25 gives access to excise tax records; and Sections 26 and 29 give access to crime and justice records. The act also stipulates that Statistics Canada is responsible for promoting the avoidance of duplication in the information collected by the various departments.

A memorandum of understanding (MOU) governs the release of administrative information to Statistics Canada. These documents say what the data are, when the data will be available, how much they will cost, and how and between whom the data will be shared. MOUs are lengthy, extremely detailed documents. For example, the MOU between the Canada Revenue Agency and Statistics Canada is over 100 pages. Creating an MOU is often difficult, involving negotiations that last years.

Another important aspect of the legal framework for linking survey data to administrative data are two policies that govern these transactions: (1) the policy on informing survey respondents and (2) the policy on record linkage.

Currently, data from different sources cannot be linked unless the Statistics Canada policy committee approves of the linkage. This committee is the highest committee at Statistics Canada, chaired by the chief statistician. However, under the policy on record linkage, two omnibus record linkage authorities have been approved and allow linkages to be performed under certain circumstances without requiring separate approval by the policy committee.

The first authority is the omnibus record linkage authority for the economic statistics program, and it allows linkage of data for business surveys. The second authority is the omnibus record linkage authority for improving population and household survey programs, which allows linking data for three reasons: (1) to improve a survey (e.g., to improve stratification, nonresponse adjustment), (2) to study and assess survey data quality (e.g., to improve survey frame quality, assess disclosure risk), and (3) to aid in data collection (e.g., to add addresses or phone numbers). Record linkage is not allowed under these omnibus authorities, however, if the purpose of the linkage is to produce estimates for public release. To do this, approval is still required from the policy committee.

Trépanier also discussed the challenges and drawbacks they experienced using administrative data. Referencing points also made by Jelke Bethlehem about the Netherlands, she commented that researchers will never have the same control over administrative data that is possible over statistical data. Even if a thorough evaluation of the administrative data is conducted before deciding to use them, there are still errors and risks that can jeopardize the process, and statistical agencies often are not informed about changes that can have these types of effects. Some of the major risks are summarized below:

- Data may change or cease to be collected without warning for some parts of the population.
- The concepts and definitions underlying data may not be exactly what is assumed or expected.
- Often quality assurance by the organization collecting the administrative data is not comparable to what could have been put into place for purposes of statistical usage.
- Timeliness of the data is frequently a problem.
- The lack of stability in the administrative data program is also a danger.

Much like the United States, Canada is encountering many challenges with household surveys. Trépanier named decreasing response rates and increasing costs as the most important. Even in the Labour Force Survey (LFS), which is mandatory, there has been a slight decline in participation. There is also a perception of an increased response burden, not only due to requests for infor-

mation from statistical agencies, but also from administrative agencies and the private sector.

Similar to the United States, Canada has considered ways of overcoming these challenges, and the use of administrative data has been identified as one option for overcoming them, because it allows for the reduction of sample size. Specifically, administrative data can be used to construct list frames, which can in turn be used to allow for stratified simple random sampling. List-type frames can make design simpler and more efficient.

Administrative data are also helpful to use in indirect estimation (calibration). Administrative data may reduce the effort required to reach each respondent, and they may be able to provide better contact information for the sampling frame. They can also be used to help implement a more efficient collection strategy, such as responsive design. Using administrative data may help reduce the volume of data collected by partially or completely replacing survey data. Furthermore, they can reduce the impact of nonresponse.

There are multiple examples of how Statistics Canada has used administrative data, Trépanier said. Even before the passage of the omnibus record linkage authority, administrative data have been used to complement existing sampling frames, such as the Address Register (AR) mentioned earlier, with additional information on addresses and telephone numbers. The AR was substituted for the listing of approximately 40 percent of clusters in the last redesign of the LFS area frame. Administrative data have also been used in the random digit dialing frame to identify a working bank of telephone numbers and to add addresses for advance letters to the residences whose telephone numbers were selected for interview.

There are also instances of using administrative data for partial substitution of other survey data. For example, rather than collecting income from respondents as part of the 2006 census and other household surveys, such as the Survey of Labour and Income Dynamics (SLID) and the Survey of Financial Security (SFS), Statistics Canada asked respondents for permission to use income tax information instead. Currently, the permission rate is about 80 percent.

Trépanier explained that Statistics Canada has used administrative data for indirect estimation in the past. Specifically, they were used to improve consistency across surveys for income estimates using harmonized calibration for the SLID, the SFS, and the Survey of Household Spending (SHS). Statistics Canada used what is referred to as T4 information, or employers' forms on salaries and wages. The number of employees by class of salaries and wages is used as a control total in the calibration in conjunction with the traditional calibration to demographic control totals. These methods were successful in improving consistency across survey estimates produced by these surveys. Administrative data have been used for direct estimates as well for tabulations of certain pension, health, justice, education, and travel statistics.

Since the passage of the 2008 data omnibus record linkage authority, an example of how administrative data have been used is to construct a frame for the new Survey of Young Canadians. Neither households rotating out of the LFS nor a fresh sample of dwellings from an area frame was sufficient or cost-effective for generating a sample for this survey. Because of the need to sample from a unique population of respondents ages 1-18, Statistics Canada turned to the Canada Child Tax Benefit (CCTB) file. Every child ages 0-6 in Canada receives a monthly benefit, irrespective of family income, and the child is registered in the hospital at birth. Children who are no longer eligible for the benefit are also included; thus the database is quite comprehensive.

In comparing the 2006 CCTB file with that of the 2006 census, it was discovered that coverage in the CCTB was quite good: 93-97 percent per age per year. Income distributions between the two collections were also quite similar. However, the Survey of Young Canadians was planned primarily as a survey using computer-assisted telephone interviewing (CATI), and contact information was not in the file received by Statistics Canada. Arrangements were subsequently made with the Canada Revenue Agency to obtain contact information, Trépanier said.

In a field test of the survey, which was mostly a test of the contact information, 83 percent of the 1,000 test cases had a valid address on the file. Also worthy of note is that there was an anticipation of concern, particularly from parents, about the use of the CCTB to reach respondents, but the pretest indicated that this was not a problem. As an example of previously described potential drawbacks of administrative data, at some point the records of all persons over age 18 were removed from the database based on the argument that they were no longer eligible for the benefit, even though they would have been of interest for the survey.

Other efforts to centralize and improve tracing operations using administrative data currently pursued by Statistics Canada include samples sent to the Canadian Council of Motor Transport Administrators (CCMTA), which returns them with addresses from driver's license information. Statistics Canada is also making greater use of the National Change of Address file that is created by Canada Post.

One recommendation put forth by the Vision for Administrative Data Task Force at Statistics Canada was to develop an explicit policy on administrative data, Trépanier said. Currently, Statistics Canada has a guiding principle for administrative data use but no official policy. In addition, centralizing processes for taking in and using administrative data need to be established, she said. This would entail creating an inventory of data and assigning management responsibility for each data source. There is also a push to mobilize existing resources, prioritize research, and establish a governance process on how to use administrative data.

For the future, Trépanier said, using administrative data to build sampling

frames is of particular interest. There is the risk of coverage error in using an administrative database in constructing a frame, but if it is done in the context of using multiple other frames and calibration to correct coverage error, this is probably less of an issue. The ideal goal is a single frame, which is the approach used in building Statistics Canada's Address Register, but this does not preclude the inclusion of auxiliary information. A single frame would allow for better coordination of samples and survey feedback, she said.

For data collection, one of the goals related to administrative data is to enable tracing. Statistics Canada wants to centralize the tracing process leading to the linking of all administrative data sources to make available the best contact information possible. This will require substantial effort, including a process to weigh the quality of the different sources and determine what contact information is most likely to be accurate. Another goal for administrative data could be to better understand the determinants of survey response and improve data collection procedures based on this information. For example, administrative data can provide guidance on preferred mode of data collection if one can assess whether persons who file their taxes electronically are also more likely to respond to an electronic questionnaire.

Statistics Canada has been successful in using substitution of income data from tax records, and this is likely to be continued. It is yet unclear, however, whether other information is available that could replace survey data. Investigating these options is done with caution because of the risk discussed. There is also the problem of ensuring consistency between survey and administrative data across variables.

Administrative data can also assist researchers in better understanding nonresponse bias and the impact of lower response rates. Finally, they can help both reduce the volume of data collected in surveys and improve estimation. Now that Statistics Canada has the omnibus record linkage authority in place, exploring all of these options has become a much easier process.

DISCUSSION

The discussion of the various methods used in the collection of household data began with several questions about the Canadian system of household surveys. Kathleen Styles (Census Bureau) asked for clarification on the omnibus record linkage authority—specifically, how did that come to pass, what was the motivation, and what did it hope to accomplish? Trépanier answered that it was established after someone realized that requests for linkage were going to the policy committee quite frequently (about every two weeks) and that many of these linkage requests were similar in nature. This process became burdensome, particularly considering that the requests generally did not involve disseminating administrative data. Since a record linkage authority already existed on the business side, that was extended for use in the area of

linking social and survey data as well. But it is important to remember that the omnibus authority was designed to be used for evaluations that could improve surveys—not to disseminate administrative data sources. And although going to the policy committee is no longer necessary, the Access Division at Statistics Canada must be notified of the administrative data use so that it can make an inventory of all the linkages.

Styles followed up her question with another one about registers. A register of persons is a loaded issue, but does Statistics Canada have permanent files that are intended to represent all Canadian residents? In the discussion of tracing and a centralized address frame, it seemed as if this may be similar to a register. Trépanier responded that the central processes for tracing are under construction now. As for the Address Register, the plan is not necessarily to use it for all of Canada. As Tambay said earlier, the AR will be good for listing in urban areas, but it is likely that there will still be a need for an area frame, particularly for rural areas.

Cynthia Clark asked Trépanier to clarify under what circumstances is Statistics Canada required to obtain consent for the use of tax data. Trépanier said that one interpretation of the Statistics Act is that permission is only necessary if administrative data were to be used in conjunction with other survey data. In those cases the respondent would need to be informed that the data are being linked.

Graham Kalton reminded the participants that according to Trépanier's presentation, the SLID obtains permission from a high proportion of respondents for the use of tax records, but about 15 percent refuse to grant permission. But researchers still have access to all the records. Is Statistics Canada now allowed to match those records together to evaluate the returns? How is this problem handled? Would it be better not to ask permission and just use the records?

Trépanier said that they were interested in conducting a study of the SLID respondents who refused access to their tax records, but it turned out that the way they are currently asking for permission is very general, and this precludes the linkage if respondents refuse.

A discussion participant asked Martinez for clarification on the integration of administrative health data, specifically, whether a linkage of the National Health and Nutrition Examination Survey (NHANES) to states is the issue under consideration or whether something more elaborate is planned. Martinez replied that, initially, the primary files being linked would be Health Interview Survey data with Centers for Medicare & Medicaid Services data, using mostly the Medicare files. The NHANES linkages to some state files are part of the other pilot study, the nutrition and food assistance project.

Jay Ryan (Bureau of Labor Statistics) is interested in new data collection technologies and asked Dillman what kind of research is being done with text messaging for survey contact, particularly now that text messaging has become

so prevalent. Also, how will the shift to larger cell phone screens, particularly in the case of smart phones and tablet PCs, affect data collection? Phillip Kott agreed that text messaging is becoming an increasingly important mode of communication among young people in particular, who often consider phone calls rude and expect a text message even before agreeing to talk to someone on the phone.

Dillman said that he was not aware of much research on text messaging, but this was something he has thought about, particularly what kind of coverage it would entail and the type of people most likely to use it. He added that he suspects that people who use text messaging frequently may be quite different from those who do not. Another concern related to this technology is that if people read text messages on the go, they are not going to stop to fill out a survey, because they are probably not in a good place to do that.

On smart phones and tablet PCs, Dillman said that the screens of many of these are still too small. Still, surveys will eventually be constructed for these devices. He predicted that the first study of surveys on smart phones and tablet PCs will happen as early as spring 2011.

This issue is a challenge even in the case of those who rely on email as their primary form of communication, Dillman continued. In the studies he has conducted of both mail and email contacts to entice survey participation, he received a higher response when a questionnaire was sent via postal mail than when an email response was requested. Young people also tend to go to paper first. The bottom line, however, is that little progress will be made on electronic surveys if all that is done is to send an email and then expect people to respond. Even for young people, surveys will need to do something different. This sometimes results in a higher cost for web surveys than mail.

Keith Rust noted that, in Westat's studies of mode choice, many respondents use more than one mode, which means that responses have to be unduplicated. This may be because respondents use a mode that is convenient to them and then use another one in addition to respond to the survey because they think that is what the administrators of the survey want them to use.

Dillman replied that it is critical that researchers be very clear about what is requested of respondents. For example, if a web response is preferred, the survey should state that and explain the reasons. Even then, giving a questionnaire to a person but then telling them to respond by another mode, web for example, is a challenge, because the respondent will consider that the paper is right there in hand and, in order to respond by web, one must wake up the computer, and type in a complex URL.

Jelke Bethlehem asked Dillman for clarification on his advice not to use CATI and computer-assisted personal interviewing (CAPI) in mixed-mode surveys but rather use mail and emails. One of the Statistics Netherlands surveys follows up web contact with mail, then CATI, and then CAPI. Does Dillman

recommend that the CATI and CAPI follow-up steps be abandoned in this survey?

Dillman clarified that he was not suggesting that any of the modes should be abandoned. Different situations call for different modes. It is, however, increasingly difficult to conduct a conversation with people over the telephone, because that is not how the telephone is used anymore. Society has evolved so that people control the phone, and they use it when they want to. It used to be that they had to answer the phone or miss a call. Changes in culture are contributing to the decline of phone surveys more than changes in technology. The technology just made the culture change possible.

5

End of Day 1:
Discussant Remarks and Floor Discussion

DISCUSSANT REMARKS

Prior to the floor discussion, Alan Zaslavsky (Harvard Medical School) summarized some of the salient points from the first day of the workshop. Referring to the series of presentations on other countries' survey systems, he noted that what is impossible to implement in one country might be the only way to do things someplace else. In the same manner, what is impossible in the United States today could be a research project in 5 years, and in 10 years it might become obvious that this once-impossible strategy is now the only way to operate. In other words, persistence can pay off.

He went on to say that the reasons for some of the differences across countries go beyond the realm of scientific considerations to areas in which participants at this workshop do not necessarily specialize: history, politics, and culture. The degree of centralization characterizing administrative structures is another important factor contributing to differences. Nevertheless, the presentations can serve as a wake-up call for the statistical community in the United States to consider household survey systems in other countries and to aspire to learn from the experience of others.

Zaslavsky mentioned that there was a lot of discussion about innovation. Now, he said, it is a question of how can the statistical system convince itself, and then others as well, that many of the ideas mentioned today are worth pursuing. In the case of the U.K. survey, validation was carried out by comparing the new series with the previous series, which from a statistical point of view is a fairly clear-cut process. But if members of the statistical system are truly interested in innovation, then they must be prepared for situations in which

55

the new measures will not be consistent with what was done before. Although changes in methodology will make some data users unhappy, a new methodology may be equally or perhaps more fit for use and more practical to implement. This may mean that agencies and decision makers will have to think hard about who the key data users are, as well as what information and policy needs have to be satisfied.

An example of a transition to a new methodology in the U.S. federal statistical system is the transportation research community's transition from using the census long-form sample to using the American Community Survey as a source of transportation data. At the start of this process, they were reportedly quite unsure about the idea of using data that were based on a rolling sample and that would usually be 2 or 3 years old, as opposed to the data from the census long-form sample, which could be up to 10 years old. This is a good example of breaking away from the way things have been done with the goal of improving the fitness for use, and now they may have something better than what they had before.

Another way of thinking about the issue of acceptability is to question what are considered official statistics. Some people argue that an actual enumeration is the only legitimate way to count the population, but the statistical community knows that this is not the best approach to obtain most of the data. The question is how far is the statistical and survey community really willing to go to innovate. When will model-based estimates be widely accepted as official statistics? There have been and continue to be challenges to almost all forms of statistical methodology applied to the census. But the statistical system is in a position that it could be releasing a lot more official numbers that are model-based, and indeed there are some areas in which model-based estimates are well accepted, such as unemployment statistics that are adjusted through a sophisticated time-series model.

There has been considerable talk of Google's consumer price index (CPI) recently. If Google develops a method that tracks the online sales of groceries, it will probably reflect the price of groceries in stores fairly well. The index will, of course, be based on a biased sample, with not nearly the right coverage of grocery stores, but if there is demand to get a leading indicator of the CPI without having to wait for data to arrive from an agency whose field representatives are visiting stores or calling people and asking what they paid for a gallon of milk, the Google CPI, or a more disaggregated version of it, can be useful for statistical modeling.

However, this does not mean that the statistical community should be accepting all new methodologies that come along. There is still an important role for statistical agencies, perhaps as gatekeepers, because raw administrative data and unvetted Internet surveys are not going to necessarily yield very good statistics.

Zaslavsky also reflected on the discussions about the use of different modes

for data collection, which may require the use of different sampling frames. There are some purposes for which Internet panels may be a useful tool—for example, they are widely used in market research. Few researchers believe that these panels are efficient, representative, or accurate as a simple statistical estimation tool. However, they are quite consistent from month to month, because respondents are on the same panel for a few years or even longer. If the research interest is to look at trends or change over time, the data from these panels may be quite useful, although only in modeling. This is another area in which the statistical community must consider how far it is willing to stretch the concept of official statistics in order to make use of tools like this.

In the day's presentations there was a good deal of discussion about the use of surveys as sampling frames for other surveys. There are obviously substantial efficiencies resulting from collaborations of this type, but there are also substantial challenges related to making these arrangements work well, Zaslavsky said. There is the problem of the second-phase survey inheriting the limitations of the first-phase survey. Beyond this, there are significant administrative barriers that exemplify many of the problems occurring in the statistical system more generally, especially different objectives that come along with different sources of funding.

Some of the important underlying issues are those of privacy and confidentiality. These concerns are very ill defined. What exactly does privacy mean? Jean-Louis Tambay gave an excellent example of how a confidentiality scandal can be created by simply informing the public of an existing data collection practice, even if there have been no breaches of confidentiality. A scandal on this topic is easy to create at any time.

One could argue that, in the past, the protection of privacy was guaranteed primarily through inefficiency and inaccessibility. For example, a great deal of public data are unalphabetized and moldering in the basements of courthouses in over 3,000 different counties. In some sense, those data are private, and it does not matter that they are actually public. Today a lot of information is easily accessible over the Internet, and as the inefficiencies are fading, organizations are finding that they must establish official policies about storing public records that were once much less obviously public. A national policy conversation is required to think about what the rational trade-offs are and the obligations of individual citizens and the polity toward each other. Zaslavsky added that it is also worth mentioning that the greatest threats to privacy and consequences of breaches are from the commercial sector, not government data collections. For example, being denied a home loan because someone stole your credit card is a scenario that is a lot more likely than confidential data being released by a government agency.

For years there has been talk of using administrative records, especially for the census, but in every case it was decided that it was not the right time. Zaslavsky has always believed that taking small steps and making incremental

progress is important to move the statistical system forward in this area. If there had been more persistent efforts in the early days, the system would be much further ahead now. Julie Trépanier presented a good list of alternative uses for administrative data and of programs actually being implemented at Statistics Canada, incremental as they may be.

Zaslavsky said that the current work in this area, described by Rochelle Martinez, is perhaps one of the most optimistic developments in years for the federal statistical system. But one question that arises in response to these initiatives is whether the opportunities for sharing will be adequate for everything that is needed. As an example, there is clearly a role for those who work with the Statistics of Income Division (SOI) to work with data from the Internal Revenue Service (IRS). The SOI can collect a sample and clean it, thus making it a much better data system than just the raw tax returns would be. These analysts can then cooperate with other agencies for data matching. However, there are some situations in which there really is a need to have access to the entire IRS database, and a statistical agency may or may not be able to gain that access. The point, Zaslavsky said, is that broader support is needed to carry out linkage projects.

FLOOR DISCUSSION

The topics covered during the floor discussion at the end of the first day were as varied as the day's presentations. Cynthia Clark commented that as part of the thinking about the sharing of sampling frames across agencies, it would be useful to consider the development of a frame that contained both households and establishments in a comprehensive geographic system. She recalled that a suggestion similar to this was made as part of the work of a United Nations commission developing a global strategy for agricultural and rural statistics. The goal of the initiative was to develop a system that enables the collection of comparable data across countries and to build a master sampling frame that would allow linkages to occur. She added that, in the National Agricultural Statistics Service, which focuses on rural statistics, access to a household sampling frame would enable the agency to better meet some of its data needs than what is currently feasible given the design of the American Community Survey.

Trivellore Raghunathan (University of Michigan) noted that, with the advent of mixed-mode designs, there needs to be an effort to understand what is really being measured, because context matters for survey participation. Research has shown that if the same question is asked in two different ways, different answers will result. Perhaps the differences should be modeled to create some sort of population-level equivalence. Jelke Bethlehem agreed, saying that in the Netherlands, much of the survey data can be collected via the web, making mixed-mode surveys cheaper. However, it is difficult to disentangle

mode effects and selection effects, and there are concerns about the estimates as a result. Developing models to examine these questions would be interesting.

Phillip Kott noted that as long as there is nonresponse in a survey, model-based methods will have to be applied. Many of the participants at the workshop recognize that models are already being used in multiple ways. For example, model-assisted methods are used to get a good sense of probability sampling properties, to carry out small-area estimation, and to create synthetic estimates. Furthermore, data users generally do not care how the data are produced; they just want them. So perhaps it is worth considering how much of the resistance to model-based estimation comes from the statistical community itself.

Roderick Little (Census Bureau) agreed that much of what is done now is model-based. The issue is the robustness of the models and how they represent the data. Regarding administrative records, he added that their role may be different depending on the intended analysis. In many cases, administrative records may be most useful for descriptive statistics, such as an income distribution, given that the records do not usually contain information about relationships.

Zaslavsky responded that in some cases it is possible to imagine administrative records being more useful for analytic purposes than survey data. An example of this would be longitudinal data, such as income tax records that go back 30 years. Survey data are rarely available for a similar time period. However, producing model-based estimates designed for descriptive purposes and then using these in analytic studies could be problematic. In an analytic study that involves a model-based estimate with a large regression component, relationships may be discovered that are primarily due to the way the model was specified. So it is important to go back to the original data and understand how they were put together in order to be able to use them in an analytic study.

Bethlehem provided an example from Statistics Netherlands to illustrate how relationships can be studied using administrative data. Statistics Netherlands combined police register data with population register data to examine relationships between ethnic background and crimes committed. He added that sometimes it is possible to study relationships that could not have been examined with survey data alone, but he acknowledged that a major limitation is that these types of data are not necessarily accessible to outside researchers because of disclosure concerns.

Frauke Kreuter (University of Maryland) said that the German Department of Labor Statistics has permission to link indicators, such as nonresponse and linkage consent indicators, to an administrative database on the grounds that they are survey production data that do not reveal personal information. This could be described as an incremental step that allows researchers to use the administrative data for modeling in various forms. It may be interesting to consider whether such a step could be within reach in the United States, she said.

Katherine Wallman said that it is time to have a conversation with the

American public about the issue of privacy. Prior to the release of the memo by the Office of Management and Budget (OMB) that outlined several pilot programs for the use of administrative records, OMB staff met with privacy advocates. Despite these conversations, it remains unclear whether many of these privacy issues have been fully parsed out with this community, and they have definitely not been parsed out with the public. She said that the federal statistical community needs to take some risks in this area and to have a carefully constructed conversation about privacy, and in her view the time to do that is now.

Wallman said that there is frequent miscommunication on the topic of administrative records, because often assumptions are made about how the data will be used without the specifics being discussed. She was reminded of this during Trépanier's very clear presentation, which made her realize that she and her Canadian colleagues have been talking past one another about the use of tax data for the past few years. She clarified that the Census Bureau does have access to tax data for most of the functions that Statistics Canada does, short of actually using the records to replace missing data. Another example recalled by Wallman involved the discussions of extending authority to the Bureau of Labor Statistics to use tax records, and this dialogue was also hindered by miscommunications related to the type of use. Wallman ended by saying that she plans to advocate for more conversations about data sharing.

6

Small-Area Estimation

FINDING THE BOUNDARIES: WHEN DO DIRECT SURVEY ESTIMATES MEET SMALL-AREA NEEDS?

This session focused on the topic of producing estimates in situations in which only a small amount of information is available or there are other limitations, such as physical, temporal, or conceptual boundaries that make direct estimation difficult. In the first presentation, Robert Fay (Westat) discussed the boundaries between direct estimation and model-based small-area estimation. He noted that model-assisted estimation (Särndal, Swensson, and Wretman, 1992) can be viewed as an intermediate point between the traditional design-based approaches and model-based estimation. For the purposes of his talk, however, he included model-assisted estimation as part of direct estimation.

Theories of design-based sampling (Neyman, 1934; Hansen, Hurwitz, and Madow, 1953), although robust and useful in many applications, are based on the asymptotic properties of large samples. However, in practice, researchers are often dealing with moderate-size samples, and even in the case of large samples, the subdomains of interest are often represented by small samples. What constitutes a sufficiently large sample size depends on the intended use of the data, and this is an important question because it determines whether direct estimation is adequate for a specific task.

Fay recalled the 1976 Survey of Income and Education (SIE), which was conducted one time only, with the goal of producing state-level estimates of children ages 5-17 in poverty, with a 10 percent coefficient of variation. Because its sample was approximately three times as large as the Current Population

Survey (CPS), the SIE generally achieved this target reliability, which was considerably better than what the CPS offered, particularly in small states. Even when reliability targets were negotiated in advance, however, the impact of sampling error on the face validity of the estimates was more pronounced than survey designers anticipated. Although the SIE met the target reliability requirements, the state of Alabama protested the large decrease in its poverty rate since the preceding census. The use of small-area models to correct data problems of this type was not yet an established practice at the time.

In some situations, model-based small-area estimation represents a necessary alternative to direct estimation. Some of the earliest examples of small-area estimation include postcensal population estimates from the Census Bureau and economic series produced by researchers at the Bureau of Economic Analysis, even though these precedents use slightly different paradigms.

Several of the basic model-based approaches to small-area estimation emerged decades ago, including:

- synthetic estimation (Gonzalez and Waksberg, 1973; Levy and French, 1977; Gonzalez and Hoza, 1978),
- area-level models (Fay and Herriot, 1979),
- structure preserving estimation (Purcell and Kish, 1980), and
- unit-level models (Battese, Harter, and Fuller, 1988).

These basic approaches were followed by a number of refinements, such as mean square estimation and hierarchical Bayes approaches, he noted. Even as these model-based approaches expanded, researchers pointed out that model-assisted estimators could represent a viable alternative in some situations (Särndal and Hidiroglou, 1989; Rao, 2003).

Fay mentioned some reviews of early applications: *Small Area Statistics: An International Symposium* (Platek et al., 1987) and *Indirect Estimators in U.S. Federal Programs* (Schaible, 1996), which was based on a 1993 Federal Committee on Statistical Methodology report and includes examples of practice from several agencies. A basic resource on theory for scholars starting out in this area is the classic *Small Area Estimation* (Rao, 2003). Another useful review of theory is *Small Area Estimation: An Appraisal* (Ghosh and Rao, 1994).

It is clear that even though the theory of model-based small-area estimation has been available for decades and a number of researchers have expanded the theory, the number of applications is not yet large. Possible reasons are that model-based estimates are more difficult to produce, replicate, and combine with other estimates than direct estimates. Model-based estimates are also more difficult to document and explain to data users. For example, even when estimates of error are produced for an annual series of small-area estimates, users are typically unable to answer other questions from the published information,

such as what reliability is achieved by averaging small-area estimates for a given area over multiple years.

Some have argued that, although experimentation with model-based estimates should be encouraged, more caution should be exercised when deciding whether to publish them because they are often not equivalent in quality to the direct estimates typically published by government agencies. One approach is to clearly distinguish them from other statistical products. Fay referred to the example of the United Kingdom and the existence of experimental versus official statistics. In the United States, model-based estimates have, in some cases, been published and endorsed. The arguments for doing so are especially strong when the data produced are more informative than any other alternatives available to users, particularly when the estimates are able to meet a legally mandated need.

One example is the Current Population Survey, including its Annual Social and Economic Supplement (ASEC). The original CPS sample design was developed with national labor force objectives in mind. The first-stage design comprised strata of self-representing primary sampling units (PSUs), mostly large counties, and non-self-representing strata, each composed of PSUs of one or more smaller counties, from which one, or sometimes two, PSUs were selected at random. Because the design was originally guided by its national objectives, non-self-representing strata typically crossed state lines.

In 1973 publication began of average annual unemployment for some of the states, accomplished by a quasi-modeling that involved collapsing strata and reweighting PSUs to compensate for effects of the national stratification. The sample was soon expanded to produce estimates of annual unemployment meeting the criterion of a 10-percent coefficient of variation in each of the states (for an underlying unemployment rate of 6 percent). To avoid continuing the quasi-modeling approach, in 1984 the design was changed to state-based stratification as part of the survey redesign, which eliminated the need for these adjustments.

The ASEC supplement in the CPS has been the source of a number of small-area applications, including the Small Area Income and Poverty Estimates (SAIPE) Program, which provides income and poverty estimates for states, counties, and school districts. SAIPE is an example of a program that fills mandated needs. Considering the small CPS sample size for the target areas, the SAIPE program was quite ambitious and largely successful (National Research Council, 2000).

After the American Community Survey (ACS) was launched, the SAIPE program moved from the CPS to the ACS because of the larger ACS sample size. The ACS includes approximately 2 million interviewed households per year. The ACS pools data over 1-, 3-, and 5-year periods to produce estimates. Although the 5-year estimates produce data for the smallest publishable geographic areas, the SAIPE program currently models the 1-year ACS estimates, most of which

are not publicly released, to increase the timeliness of the data, and relies on small-area models in place of averaging over time to reduce the relative impact of sampling error. It will be interesting to assess the trade-offs related to the different releases after two or three sets of 5-year ACS estimates become available. It will also be of interest to observe whether the sampling variability of the ACS will encourage a new series of small-area applications to replace the ACS direct estimates for some uses.

Fay mentioned another case study that is worth following closely: Canada's National Household Survey (NHS), which replaces the mandatory long questionnaire that one in five households used to receive as part of the Canadian population census. Although details are still emerging, the current plans for a voluntary survey partially integrated into the census are likely to result in lower response rates and higher variances compared with previous censuses. The case of the NHS could become an unplanned experiment in what happens when data become less reliable than users have grown to expect.

Fay also briefly mentioned his work as part of a Westat team commissioned by the National Center for Health Statistics to evaluate options for averaging several years of data from the National Health and Nutrition Examination Survey (NHANES) to produce quasi-design-based estimates for the state of California. The products will be both a weighted file and an approach to estimate the variance of the estimates.

Looking ahead, Fay predicted that the area on the boundary between traditional design-based survey estimates and small-area estimates will probably grow in importance because there is an increasing demand for subnational estimates, surveys costs are rising, and modeling tools represent a possible route for incorporating existing administrative records into the estimates. Review of the case studies presented and similar ones can help guide the evolution of policy on the use of small-area estimation at federal statistical agencies.

USING SURVEY, CENSUS, AND ADMINISTRATIVE RECORDS DATA IN SMALL-AREA ESTIMATION

William Bell (Census Bureau) discussed strategies of combining data from several sources—sample survey, census, and administrative records—to produce small-area estimates. To illustrate these procedures, he used examples from the Census Bureau's Small Area Income and Poverty Estimates Program, which combines data from different sources to provide income and poverty estimates that are more current than census information for states, counties, and school districts. Specifically, SAIPE relies on

- direct poverty estimates from the ACS (and previously the CPS),
- prior census long-form sample poverty estimates,
- tax data from the Internal Revenue Service (IRS),

- information from Supplemental Nutrition Assistance Program (SNAP) records, and
- demographic population estimates.

All data sources, including the ones used for SAIPE, are subject to various types of error, and these must be taken into consideration when making decisions about how the data can best be combined. Bell mentioned some of the main types of error affecting data sources:

- sampling error (the difference between the estimate from the sample and what would be obtained from a complete enumeration done the same way),
- nonsampling error (the difference between what would be obtained from a complete enumeration and the population characteristic of interest), and
- target error (the difference between what a data source is estimating—its target—and the desired target).

Table 6-1 shows error types most likely to affect survey, census, and administrative records data, although all three data sources could include all three types of errors. "Census" data may or may not have sampling error, depending on whether they refer to the complete enumeration or to data from a prior census sample (also known as the census long form). The distinction is important for modeling purposes.

Both the ACS and the CPS provide data suitable to produce estimates of poverty, albeit in slightly different ways, Bell said. Their weaknesses are that they are subject to large sampling error for small areas, particularly the CPS. The census estimates have negligible (state level) or low (most counties) sampling error, but the estimates become gradually more outdated after the census income year, which is essentially a form of target error. For administrative records, sampling error is usually not a concern, but the data are subject to nonsampling error, and they are not collected with the specific goal of measuring poverty, which leads to a form of target error (for SAIPE's purposes). In particular, the IRS tax data leave out many low-income people who do not need

TABLE 6-1 Typical Sources of Error for Different Data Sources

Data Source	Error		
	Sampling	Nonsampling	Target
Sample survey	X	X	
Census	Maybe	X	X
Administrative records		X	X

SOURCE: Workshop presentation by William Bell.

to file tax returns, while in the case of SNAP, the qualifications for the program are different from the determination of poverty status and not everyone who is eligible participates.

Taking into consideration the errors described, there are different options for combining these data sources, Bell said. Suppose y_i is a survey estimate of desired population quantity Y_i for area i, and z_i is a related quantity from another, independent data source. The question is how to combine y_i and z_i to produce an improved estimator of Y_i.

One option for combining the data sources is to take $\phi_i y_i + (1 - \phi_i)z_i$ with $\phi_i \propto Var(y_i)^{-1}$ and $1 - \phi_i \propto Var(z_i)^{-1}$. This assumes that the estimates from the two surveys, y_i and z_i, are both unbiased estimates of the target, Y_i, which rarely happens in practice.

An alternative is to take a weighted average of the estimates with weights instead proportional to the mean squared errors (MSEs): $\phi_i y_i + (1 - \phi_i)z_i$ with $\phi_i \propto MSE(y_i)^{-1}$ and $1 - \phi_i \propto MSE(z_i)^{-1}$. This assumes that the mean squared errors are known, or equivalently the biases are known, which is also rare in practice.

One could take one of these estimates, y_i, and use it to define the target—in other words, assume that it is unbiased. One can then use ordinary least squares regression to adjust z_i to provide an unbiased predictor of $Y_i : \hat{Y}_i^{syn} = \hat{\alpha}_{OLS} + \hat{\beta}_{OLS} z_i$, where syn indicates a synthetic estimator.

For a more formal modeling approach (Fay and Herriot, 1979), the following structure is assumed:

$$y_i = Y_i + e_i$$
$$= (x_i' \beta + u_i) + e_i$$

where:

y_i = direct survey estimate of population target Y_i for area i

e_i = sampling errors that are assumed to be independently distributed with a normal $N(0, v_i)$ distribution, with v_i assumed known

x_i = vector of regression variables for area i

β = vector of regression parameters

u_i = area i random effects (model errors), which are assumed to be independent and identically distributed according to $N(0, \sigma_u^2)$ and independent of e_i.

To illustrate this with an example from SAIPE, Bell discussed the state poverty rate model for children ages 5-17. The direct survey estimates, y_i, were originally from the CPS (three-year averages) but have since been replaced with single-year estimates from the ACS. The regression variables for each state include a constant or intercept term and

- a "pseudo poverty rate" for children, calculated based on the adjusted gross income and the number of child exemptions on the tax return;

- the tax "nonfiler" rate, which is the percentage of the population not represented on tax returns;
- the SNAP participation rate, which is the number of participants in the program divided by the population estimate; and
- census data in one of two forms, either the estimated state poverty rate for school-age children ages 5-17, or residuals from regressing previous census poverty estimates for ages 5-17 on other elements of xi for the census year.

One generally has reasonable estimates of the sampling variances, v_i. If one also had estimates of σ_u^2, their sum would provide an estimate of the variances of the y_i. Since the various sampling errors and random effects are independent, the estimated covariance matrix for the y_i is $\sum = diag(\sigma_u^2 + v_i)$, with the off-diagonal terms equal to zero given the assumed independence. Using this covariance matrix, we could estimate β using weighted least squares as follows:

$$\hat{\beta} = (X' \Sigma X)^{-1} X' \Sigma^{-1} y,$$

where $y = (y_1, \ldots, y_m)'$ and X is m x r with rows x_i'.

Turning things around, given the v_i and some initial estimates of β, one could estimate σ_u^2 using the method of moments, maximum likelihood estimation, REML, or through a Bayesian approach. (One might iterate from an initial estimate of β by setting the σ_u^2 equal to some initial value.)

It would then be possible to combine the direct survey estimates and the regression estimates using the best linear unbiased prediction (BLUP) as follows:

$$\hat{Y}_i = b_i y_i + (1 - b_i) x_i' \hat{\beta}$$

where $b_i = \sigma_u^2 / (\sigma_u^2 + v_i)$.

Bell said that a way to think about how the data are being used for small-area modeling and prediction assumes that there is a regression function that describes the variation of the mean in the target from state to state as a function of x_i: $E(Y_i) = E(y_i) = x_i' \beta$. It then follows that $E(x_i'\hat{\beta}) = x_i'\beta = E(Y_i)$ so the fitted regression can be thought of as a predictor of the target Y_i. For example, if there is only one regression variable, z_i, plus an intercept, then $x_i'\hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 z_i$. The model fitting makes a linear adjustment to the data source z_i, which otherwise has target error. After the adjustment, the fitted linear function of z_i can be used to better predict Y_i.

The BLUP is the weighted average of two predictors of the target, the

direct estimate and the regression fit, where the weights are inversely proportional to the variances of the errors in the two predictors, that is,

$$b_i = \sigma_u^2 / (\sigma_u^2 + v_i) \propto 1 / v_i;\ 1 - b_i = v_i / (\sigma_u^2 + v_i) \propto 1 / \sigma_u^2.$$

To illustrate the improvements in accuracy resulting from this modeling, Bell compared the approach of regressing the CPS poverty rate for children ages 5-17 on the pseudo poverty rate from tax records with the Fay-Herriot model, with one regression variable (the pseudo poverty rate), and with the SAIPE production model that brings in other regression variables.

Using data from 2004, let y_i = CPS 5-17 poverty rates and z_i = pseudo poverty rate for children. Regressing y_i on z_i using ordinary least squares gives the synthetic predictor

$$\hat{Y}_i^{syn} = \hat{\alpha}_{OLS} + \hat{\beta}_{OLS} z_i$$
$$= -.18 + .82 z_i$$

The analogous Fay-Herriot model is $y_i = Y_i + e_i$. In contrast to the OLS model, here weighted least squares are used, weighting inversely proportional to $Var(y_i)$, to estimate the regression coefficients. Then the regression estimates are combined with the direct estimates, weighting the regression estimates inversely proportional to σ_u^2 and the direct estimates inversely proportional to v_i.

Table 6-2 shows the mean squared errors of the two predictors for four states. The synthetic predictor is worse than the direct estimate, except in the case of Mississippi. The mean squared errors for the Fay-Herriot model with one regressor are lower than the variances of the direct estimates. The improvement is larger in the states with smaller samples and higher sampling variances, as is typical in this context. The last column in the table shows the weights that are applied to the direct estimate. For example, in California, approximately 80 percent of the weight is for the direct estimate—in other words, the model prediction is going to be very close to the direct estimate in this state.

TABLE 6-2 Prediction Mean Squared Errors (MSE) for 2004 Poverty Rates for Children Ages 5-17 Based on the Current Population Survey Target and the Fay-Herriot Model with One Regressor (FH1)

State	n_i	v_i	MSE $(\hat{Y}_i^{syn})$	MSE $(Y_i \| y, FH1)$	b_i
California	5,834	1.1	7.7	.9	.80
North Carolina	1,274	4.6	4.7	2.3	.50
Indiana	904	8.1	9.0	3.4	.36
Mississippi	755	12.0	6.3	3.9	.26

SOURCE: Workshop presentation by William Bell.

TABLE 6-3 Prediction Mean Squared Errors (MSE) from the Fay-Herriot Model with One Regressor Compared to Those of the Full SAIPE Production Model

State	v_i	MSE(Y_i|y, FH1)	MSE(Y_i|y, full model)	b_i
California	1.1	.9	.8	.61
North Carolina	4.6	2.3	2.0	.28
Indiana	8.1	3.4	2.0	.18
Mississippi	12.0	3.9	2.9	.13

SOURCE: Workshop presentation by William Bell.

Table 6-3 compares the MSEs of the one-regressor Fay-Herriot model to the MSEs for the full SAIPE production model. The mean squared errors are lower with the full model, and, again, the difference is bigger in the case of smaller states, where the predictions are less able to rely on the direct estimates.

Bell also discussed an extension of the Fay-Herriot model to a bivariate version, which can be used for modeling two statistics simultaneously. The targets in the two equations are different in this case, and there are procedures for model fitting and prediction that can potentially improve the estimates for both quantities. The bivariate model is written

$$y_{1i} = Y_{1i} + e_{1i} = \left(x_{1i}'\beta_1 + u_{1i}\right) + e_{1i}$$
$$y_{2i} = Y_{2i} + e_{2i} = \left(x_{2i}'\beta_2 + u_{2i}\right) + e_{2i}$$

This approach is useful when there are estimates of ostensibly the same characteristic from two independent surveys, such as the state poverty rates for the 5-17 age group from the CPS and the ACS. It can also be used for estimates of two related characteristics from one survey, such as the poverty rates for the 5-17 and 18-64 age groups from the CPS, or for estimates of the same characteristic but for two time points, such as poverty rates for the 5-17 age group from this year and last year's CPS.

In cases in which there are two estimates of ostensibly the same characteristic from two surveys, researchers have to decide which of the two estimates defines the target (as being the expectation of one of the estimates). One way to think about this is to consider which of the two surveys is suspected of having lower nonsampling error. If both estimates are thought to have similar levels of nonsampling error, then one may let the direct estimate that has lower sampling variance define the target, and to try to improve that estimate by modeling.

If one of the two estimates has some sort of "official" status, then this estimate could define the target. In any case, the bivariate model will utilize

the regression variables and the estimates from the other survey to predict the specified target. This adjusts for bias due to differential nonsampling and target error between the two survey estimates, but it does not address any bias in the survey estimate that is used to define the target.

The approach to generating the SAIPE income and poverty estimates is fairly unusual in the federal statistical system. Yet the estimates are widely used for administering federal programs and allocating federal funds. In Bell's view, there were several factors that contributed to the acceptance of the model-based estimates among data users. First, the modeling relies on high-quality data sources that can generate good-quality estimates. Second, the time was right for this initiative when the Improving America's Schools Act was passed in 1994, requiring the allocation of Title 1 education funds according to updated poverty estimates for school districts for the 5-17 age group, unless the model-based estimates were deemed "inappropriate or unreliable." In addition, a panel of the Committee on National Statistics that reviewed SAIPE methods and initial results also recommended that the model-based estimates be used (National Research Council, 2000).

ROLE OF STATISTICAL MODELS IN FEDERAL SURVEYS: SMALL-AREA ESTIMATION AND OTHER PROBLEMS

Trivellore Raghunathan (University of Michigan) discussed research areas in which model-based estimation represents an ideal tool that allows researchers to use data for purposes beyond what they were intended for. He noted that there has been a recent increase in the complexity of research conducted using data from federal surveys. The data available from a single survey often do not meet these complex research needs, and the answer is often a model-based approach that can synthesize and integrate data from several surveys. Some of the arguments for combining data sources include

- extending the coverage,
- extending the measurement,
- correcting for nonresponse bias,
- correcting for measurement error, and
- improving precision.

Raghunathan presented four examples from his own work. The first one involved combining estimates from the National Health Interview Survey (NHIS) and the National Nursing Homes Survey (NNHS), with the goal of improving coverage. The variables of interest were chronic disease conditions. Data were available from both surveys for 1985, 1995, and 1997. The initial strategy was a design-based approach, treating the two surveys as strata. Current work involves Bayesian hierarchical modeling to obtain subdomain

estimates for analysis of health disparity issues based on education and race (Schenker et al., 2002; Schenker and Raghunathan, 2007).

Another project involved matching respondents from the NHIS and the NHANES on common covariates involving a propensity score technique. The NHIS collects data about disease conditions in a self-reported format, which raises concerns about underreporting due to undiagnosed disease conditions, especially among those least likely to have access to medical care. The NHANES has a self-report component, but it also collects health measurements. This allowed the researchers to model the relationship between the self-reported data and clinical measures in the NHANES and then to impute "no" responses to questions about disease conditions in the NHIS using the model from the NHANES. After applying this correction to the NHIS, many of the relationships "became more reasonable." Current work focuses on extending the approach to several years of data and on obtaining small-area estimates of undiagnosed diabetes and hypertension (Schenker, Raghunathan, and Bondarenko, 2010).

The third project combined data from two surveys with the goal of providing small-area estimates of cancer risk factors and screening rates to the National Cancer Institute (NCI). In the past, NCI has relied on the Behavioral Risk Factor Surveillance System (BRFSS) to construct these estimates. However, the BRFSS is a telephone survey that faces increasing challenges associated with uneven landline coverage and low response rates. Raghunathan and his colleagues combined the BRFSS data with data from the NHIS, which covers both telephone and nontelephone households and has higher response rates.

The technique selected for this study was a hierarchical model, treating NHIS data as unbiased estimates and BRFSS data as potentially biased estimates. These assumptions were made because of the face-to-face mode and higher response rates in the case of the NHIS. The telephone household estimates from the NHIS and the telephone household estimates from the BRFSS were used to correct for nonresponse bias associated with the nontelephone households and then produce a model-based estimate for all counties.

Although in the past the concept of nontelephone households was understood to mean households without a telephone, it is becoming increasingly important to distinguish between households that do not have a telephone at all and households that do not have a landline but do have a cell phone, because the demographic characteristics of these two types of households are different. The model thus becomes a four-variate model.

Raghunathan mentioned that although the NHIS and the BRFSS are both surveys of the Centers for Disease Control and Prevention, accomplishing the data sharing still involved substantial work. A predecessor to this project, which involved linking data from the National Crime Victimization Survey and the Uniform Crime Reporting Survey, also experienced challenges related to confidentiality and privacy restrictions. Raghunathan emphasized that these are

issues with which the federal statistical system will have to grapple in research of this type.

Raghunathan's current project involves developing an alternative to the current National Health Expenditure Accounts. The goal is to study the relationship between health care expenditures and population health, with a focus on specific elements, such as disease prevalence and trends; treatment, intervention, and prevention effectiveness; and mortality, quality of life, and other health outcomes. The relationships are examined using Bayesian network modeling, and microsimulations are performed to evaluate hypothetical alternative scenarios.

Given that no existing data set contains all of the desired measures, Raghunathan and his colleagues are working on combining data from a variety of sources. For example, the team identified 120 disease categories with major impact on expenditures. Related data for subsets of diseases are available from

- Self-report sources: NHIS, NHANES, Medical Expenditure Panel Survey (MEPS), Medicare Current Beneficiary Survey (MCBS),
- Clinical measures: NHANES, and
- Claims: MEPS, MCBS.

Although information on past and current disease conditions is available from self-report data, the claims data represent current conditions, so to combine the information, both types of data are converted to a measure of whether the person ever had the disease. For example, Figure 6-1 shows the information available from the MCBS and the NHANES. Respondents are matched on the covariates and then the missing self-report in the MCBS is imputed, so that the overall self-report rates in the two surveys agree.
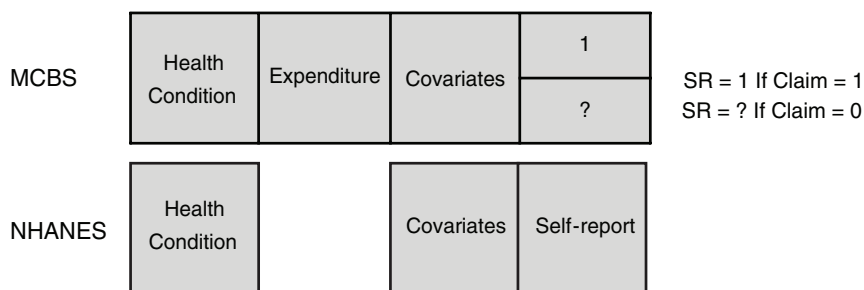
FIGURE 6-1 Data layout for the Medicare Current Beneficiary Survey (MCBS) and the National Health and Nutrition Examination Survey (NHANES).
SOURCE: Workshop presentation by Trivellore Raghunathan.

Raghunathan concluded by saying that although there are a lot of challenges related to the portability of information from one survey to another, including differences in the data collection instruments, protocols, and timing, often combining several data sets is the best option. When data from multiple sources are synthesized, statistical modeling and an imputation framework are particularly useful tools to create the necessary infrastructure. However, data sharing has to be made easier for these types of approaches to be successful. In an ideal world, all data collection would be part of a large matrix, with a unified sampling frame and complementary content that could be filled in by different agencies or researchers working on different pieces.

DISCUSSION

Roderick Little started the discussion by saying that the term "design-based" theory of sampling conflates the design aspects of the work with the analysis aspects, and that it is perhaps more appropriate to think of it as design-based theory of inference. Little described himself as a modeler and an advocate of what he calls "calibrated Bayes," in which the design and the repeated sampling aspects come into the assessment of the model rather than into the actual inference (Little, 2006). This approach makes it possible to avoid "inferential schizophrenia" between being design-based or model-based. He prefers to think of everything as models, and, in that sense, the design-based model can be described as a saturated fixed-effects model, in which one does not make strong structural assumptions, so there are no random effects. One can also consider unsaturated hierarchical models, so to the extent that there are any distinctions, they are in terms of the different types of models.

Little argued that hierarchical models are the right way to think about this conceptually. The advantage of hierarchical models is that it is not necessary to use either the direct estimates or the model-based estimates, because they provide a compromise between the direct estimate from the saturated model and the model-based estimate from the unsaturated model. Fay's discussion of SAIPE illustrates how it is possible to get estimates for different areas, some borrowed mostly from direct estimates and some from the model-based estimates.

In some cases, Bayes modeling may be a little better because it does a better job with the uncertainty and the variance components. While the calibrated approach is a weighted combination of the two, the weights can be poorly estimated, and in simulation studies the calibrated approach can provide worse inference than the model-based approach when the hierarchical model is reasonable. Little finished by stating that the challenge is to come up with some kind of index of what he called structural assumption dependence. For example, when the weights allocated to the model get too high, it might be possible to use that as a criteria for whether to publish an estimate.

Other aspects of this include how well the model is predicting and the extent to which the model is highly parametric. He said that research is needed to develop the right indices.

Fay responded that he will have to think about some of Little's points, but that he wanted to defend the need for a boundary because it is a practical tool for statistical agencies. The number of people who can implement the design-based theory of inference is much larger than those with the skills described by Little, so that represents a very practical boundary. Identifying the boundary will help those who have to decide whether they want to pursue a small-area application that requires considerable effort and buildup of staff. In response, Little responded that, since this is a forward-looking workshop, the emphasis is not on how things are now, but on thinking about how things might be in the future.

Graham Kalton asked Raghunathan whether using Medicare administrative records was considered when producing estimates about the population ages 65 and older. Raghunathan responded that he is a "scavenger for information," using as much data as he can find, and he did explore the Medicare claims information, which is now part of the administrative data used for the fourth project he discussed. He agreed that the quality of the auxiliary data is very important in order to borrow strength for small-area estimation. In his third project, he and his team worked hard on obtaining county-level data from a wide variety of sources, not only census data, but also marketing data and data about how active the public health department is.

He added that they also went through a lot of soul searching in terms of whether the estimates are publishable. They had a meeting at the Centers for Disease Control and Prevention with state health department representatives and presented the estimates. Most said that the numbers looked reasonable. The few who did not, also came around after they went back and thought about it. The fact is that researchers have to rely on the best information available to solve a particular problem, and the modeling framework provides an opportunity to move forward with research on these topics.

Raghunathan commented that in some areas of statistics modeling is widely used, but the techniques are less common in the survey field. He argued that the distinctions made by survey researchers between model-based, model-assisted, and design-based approaches are not particularly helpful. In his research, they relied on the latest computational and statistical developments in the field as a whole, and that allowed them to solve the problems at hand. Quoting George Box, he said that all models are wrong, but some are useful. Viewing models as a succinct summary of the information available and using that to make projections about the population helps scientific areas, such as health policy, move forward.

Regarding Fay's presentation, Kalton commented that state stratification makes a lot of difference if state-level small-area estimates are of interest, as

they were in the California case discussed. A related issue is the number of sampled PSUs in each small area; if there is not a sizable number of PSUs in an area, direct variance estimates will be too imprecise, leading to the need to model the variances.

Fay responded that the problem of degrees of freedom raised was a common one. The NHANES has certainly lived with few degrees of freedom before. In the case of the eight years of data in California, about half of the PSUs were self-representing, which means a lot of degrees of freedom for that half of the population. The study did poorly in the remaining part of the state. He agreed that a distinction can be made between design-based estimation and design-based inference, adding that the variances may have to proceed out of some form of generalization. This was true for the CPS case as well, because for the averages it was only a guess what the true variances were.

Kalton quoted Gordon Brackstone, former assistant chief statistician at Statistics Canada, who many years ago said that the problem with small-area estimates is that the local people know better, and they will always challenge model-based estimates that appear wrong. Kalton said that it turns out that the local people do not necessarily know better, and that surprisingly they tend to react to the estimates presented by constructing a rationalization for why the estimates are correct. At least early on, there were not a lot of challenges to the SAIPE estimates.

Bell said that he believes that when Kalton spoke of large errors, he was referring to the school district estimates and also some counties. The issue was the paucity of the data they had at the school district level. In the case of the estimates that the panel chaired by Kalton was reviewing (National Research Council, 2000), the updates were coming from the county level, and there were no school district level updates, so the quality of those estimates was not as good as the data that were available for higher levels of geography. The general principle is that the smaller the area, the worse the data are going to be, and that is an issue. In recent years, SAIPE has also brought in IRS data, but the program is not always able to assign the tax returns to school districts.

Regarding challenges, Bell commented that they are continuing to get challenges, although he does not deal with them directly himself. Often they come from very small school districts, where it is easier for the locals to have a good sense of what is going on. Occasionally the challenges make reference to other data, such as free and reduced price lunch data, a situation that indicates that there is some confusion, given that these are not the same as the poverty estimates. There were also a lot of challenges based on the 2000 census data, using the census numbers to estimate the school district to county shares of poverty and making reference to what the previous estimates were. Generally, data users compare the current estimates to something else, and they tend to react when they see a large discrepancy, even though it is clearly possible that the other estimate was incorrect. Sometimes they have a legitimate case and it

is clear that the estimates are far enough out of line, but Bell and his team are not correcting for statistical errors.

Little referred back to Fay and Raghunathan's points about the skills needed to conduct these types of analysis, arguing that it does not help to think about survey sampling as a field separate from the general statistical community in which models are being taught. Zaslavsky added that if the general feeling is that there are not enough people who can do this type of analysis, then it is important to think about the implications for new directions in training.

Fay said that this debate has been going on for many years, and the concern about model-based estimation has always been that data users cannot understand the complex techniques and are suspicious of what is going on "behind the curtain." But if data users really understood what is involved with design-based estimation, for example, postsurvey adjustment and variance estimation, they would be concerned about that as well.

He thinks it would be useful for researchers to continue to pursue this research and talk to the data users in contexts similar to that described by Raghunathan. To the extent that researchers are able to communicate their methods and demonstrate a commitment to accuracy, it is likely that data users will embrace these techniques, in the same way they accepted the classical estimators that they do not fully understand.

7

Survey Content

PROMOTING CONSISTENCY:
THE CASE OF DISABILITY MEASURES

Most federal surveys come through the Office of Management and Budget's (OMB) Office of Information Regulatory Affairs (OIRA) as part of the review process required by the Paperwork Reduction Act. Margo Schwab, of OIRA's Statistical and Science Policy Branch, described the work of her office in promoting consistency among federal surveys in the area of disability measures.

One reason for promoting consistency is to reduce the confusion among the public and Congress caused by the availability of different statistics on what appear to be similar or identical concepts. The questions generated by measures of the number of people without health insurance in the context of the health care debate are a case in point. The Patient Protection and Affordable Care Act calls for consistent measures on disability to assess disparities, and OMB has been helping agencies realize this goal. Although consistency is not always possible (e.g., several federal agencies have their own definitions of disability that determine participation in various benefit programs), the availability of a substantial body of research on the disability measures provides a strong foundation for developing consistency in areas in which there are no statutory constraints.

Another motive for pursuing consistency is the hope that coordinated measures will allow researchers to gain deeper insight into various dimensions of disability and related policies—for example, such topics as health, housing, and transportation, which are often measured separately by different agencies. Finally, disability measures are well suited for a modular data collection plat-

77

form, allowing for the flexible gathering of additional information of interest to researchers, along with the disability data.

Although the definition of disability is context dependent, in recent years the concept has shifted from a focus on physical condition, disease, and impairment to more emphasis on functional limitations caused by these factors. This involves measuring limitations and outcomes separately to understand how disparities in outcomes may be eliminated. For example, the Bureau of Labor Statistics assesses disability independently and then reports employment outcomes.

The work underlying the development of the conceptual framework for measuring disability that OMB now supports for most federal surveys was initiated in the context of the American Community Survey (ACS). The Census Bureau assembled an interagency group, which included, among others, researchers from the Veterans Administration, the Department of Housing and Urban Development, the National Science Foundation, the Department of Education, the Bureau of Labor Statistics, and the Bureau of Justice Statistics, as well as various agencies in the Department of Health and Human Services. The group reviewed the legislative mandates and needs for disability data in the context of various programs and evaluated the restrictions imposed by the format of the existing questions on the ACS. The primary measurement objective identified by the group was what Schwab called "equalization of opportunity;" in other words, a measure that could identify persons who are at risk of discrimination or who lack adequate opportunities for participation in social life as a result of their limitations in functioning. Another goal was to measure severe disability in order to identify persons who need assistance to maintain independence.

Box 7-1 shows the new disability measures used in the ACS. The questions cover limitations in vision, hearing, mobility, cognitive functioning, and self-care. Those over 15 years of age are also asked about their ability to interact with their environment, including their ability to do errands alone.

The measures developed for the ACS are now used on a variety of government surveys, including the Current Population Survey (CPS), the National Crime Victimization Survey (NCVS), the National Health Interview Survey (NHIS), the Survey of Income and Program Participation (SIPP), and the American Housing Survey (AHS).

A key characteristic of the measures is the modular platform that allows various agencies to combine the items with additional questions of particular interest to their work. For example, transportation researchers can add questions about mobility, and surveys focused on employment can add questions about accommodations in the workplace. Using the same set of key measures across a variety of studies will allow researchers to examine different dimensions of disability, and they are just beginning to reap the benefits.

Schwab said that OMB endorses wider use of the measures because they are the result of a thorough review of the existing literature and extensive test-

---

BOX 7-1
American Community Survey Disability Measures

For sample persons 1 year of age and older:
1. Is this person deaf or does he/she have serious difficulty hearing?
2. Is this person blind or does he/she have serious difficulty seeing even when wearing glasses?

For sample persons 5 years of age and older:
3. Because of a physical, mental, or emotional condition, does this person have serious difficulty concentrating, remembering, or making decisions?
4. Does this person have serious difficulty walking or climbing stairs?
5. Does this person have difficulty dressing or bathing?

For sample persons 15 years of age and older:
6. Because of a physical, mental, or emotional condition, does this person have difficulty doing errands alone, such as visiting a doctor's office or shopping?

SOURCE: Workshop presentation by Margo Schwab based on U.S. Census Bureau (2010).

---

ing, including cognitive testing and focus groups conducted by the National Center for Health Statistics. A reliability study with a split sample was also performed as part of the 2006 fielding of the ACS. This does not mean, however, that the measures could not be further improved. There is indication that using a severity scale for each of the questions about limitations might be more useful than the current dichotomous (yes/no) answer options. A measure of upper body mobility could also be added. Research is also ongoing to understand how much of the difference in prevalence estimates is due to differences in the way the surveys using these measures are administered, particularly differences in survey mode, the purpose of the survey, and whether the respondent is answering a question about himself or herself rather than about others in the household.

DIFFERENT MEASURES FOR DIFFERENT PURPOSES: THE CASE OF INCOME AND POVERTY MEASURES

Charles Nelson (Census Bureau) discussed income and poverty estimates produced by the Census Bureau to illustrate situations in which a variety of complementary measures may be most appropriate. His talk focused on the estimates from the Annual Social and Economic Supplement (ASEC) of the Current Population Survey and from the American Community Survey, two

surveys that conceptualize income in the same way but are methodologically different. Other sources of income and poverty data include the Survey of Income and Program Participation and the Small Area Income and Poverty Estimates (SAIPE) Program.

The CPS ASEC, considered the source of official poverty estimates in the United States, is a computer-assisted telephone interview conducted in the spring of every year. It collects information about the previous year's income based on a detailed set of questions that cover approximately 50 possible sources of income. It also collects information about benefits, including non-cash benefits, for a broader picture of economic well-being. The CPS has also been the source of key statistics on related topics, such as employment, Nelson said. One of its strengths is the flexibility of its content relative to that of the ACS, while maintaining the continuity of the measures in the core topic areas.

The ACS is conducted as a mailout/mailback survey throughout the year, and the reference period for the income questions is the previous 12 months. The "annual estimate" from the ACS thus becomes an estimate that spans two years, depending on when the survey is conducted. The number of questions on the topic of income is smaller than in the CPS, with approximately eight broad questions covering all sources of income. The strength of the ACS is that it produces data for all levels of geography on a wide range of topics.

The Census Bureau is planning on releasing new poverty estimates, called the supplemental poverty measure (SPM), which is based on recommendations of a panel of the Committee on National Statistics (National Research Council, 1995) and broadens the concept of income beyond money income. The Census Bureau has been researching alternative poverty measures for the past few years, and content flexibility in existing surveys has become an increasingly relevant issue as part of this work. The CPS is able to capture more benefits (e.g., school lunch, rental subsidies) and additional expenses (e.g., out-of-pocket medical expenses, child support) than the ACS.

According to Nelson, the Census Bureau's data release schedule reflects the challenges related to publishing different estimates on income and poverty from the two sources. During the early years of the ACS, its numbers were always released before the CPS, and the strategy was to highlight topic areas other than income and poverty as part of the ACS release. But in 2003 the income and poverty numbers from the ACS nevertheless received substantial attention in the media, and this created a lot of confusion when the CPS data were released a week or two later. In subsequent years, the ACS and the CPS income and poverty estimates were released together in a single press conference. During the past two years, the Census Bureau has started releasing the CPS numbers first.

An exact match analysis revealed no systematic differences at the national level between the two data sources. Figure 7-1 shows the poverty estimates, and Figure 7-2 shows the median household income from the two surveys for
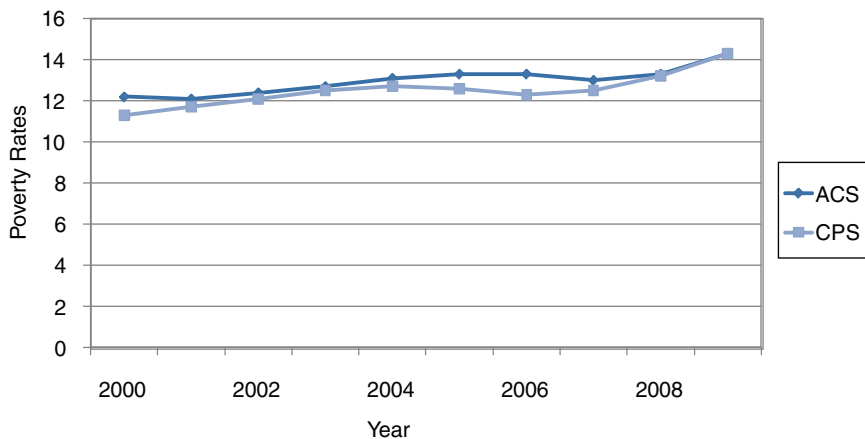
FIGURE 7-1 Current Population Survey and American Community Survey poverty rates, 2000-2009.
SOURCE: Workshop presentation by Charles Nelson.
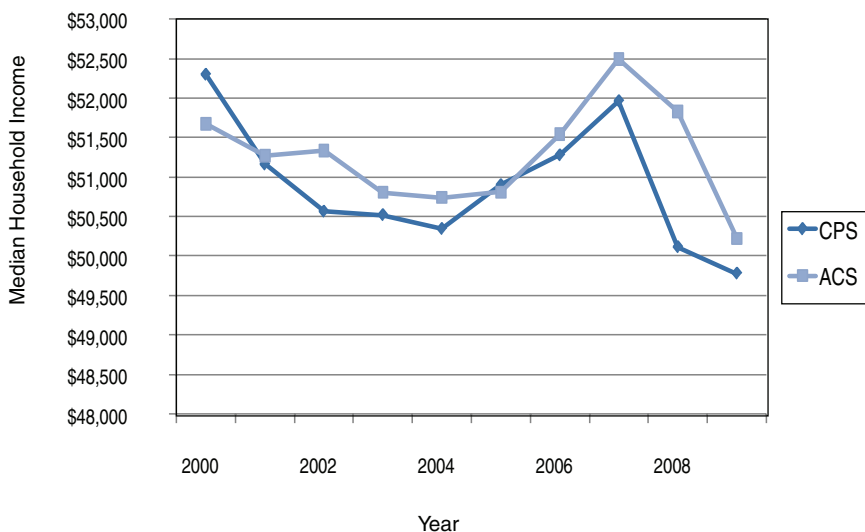
FIGURE 7-2 Current Population Survey and American Community Survey median household income, 2000-2009.
SOURCE: Workshop presentation by Charles Nelson.

the period 2000-2009. The estimates are remarkably similar, considering the differing methodologies.

At the state level, the differences are more noticeable. Comparing two-year (2008-2009) CPS averages to 2009 ACS estimates shows overall strong correlations between the two sets of state estimates. However, 24 states and the District of Columbia had statistically different median household incomes, and 16 states had statistically different poverty rates. In most of these states, the ACS poverty rate was higher than the CPS one.

Although over the long run survey sponsors and data users may be helped by complementary measures such as these, in the short run it is often difficult to explain the differences. Nelson cited the example of the headline "Census Data Give Contradictory Views on State of Child Poverty in Maryland," published in the *Baltimore Sun*, referring to a poverty rate of 7 percent based on the CPS and 13 percent based on the ACS.

In addition to being prepared to answer questions about discrepant estimates, the Census Bureau's current strategy is to minimize the overlap in the release schedules, and highlight the strengths of each of the data sources, such as the national-level time-series data for the CPS and the subnational estimates for the ACS. The Census Bureau has also been placing emphasis on releasing documentation about the surveys' methodology concurrently with the data in order to assist data users in interpreting the numbers. This includes a fact sheet summarizing the differences between the CPS and the ACS income and poverty estimates, a guide for when to use the ACS and when to use the CPS income data, and information about additional sources of income and poverty estimates produced by the Census Bureau.

In terms of the timing of the releases, Nelson said, the Census Bureau's experience shows that there is no easy answer, because data users, especially the media, are more likely to use the data that are released first. During the years when the ACS was released before the CPS, there was a tendency for people to use the ACS as the "official" poverty and income estimates, rather than wait for the CPS. When the ACS and the CPS were released simultaneously, media coverage often mixed the two sources of data, and there were also logistical issues to overcome in coordinating the schedule of the two surveys. During recent years, when the CPS was released first, users often turned to the CPS for subnational data, even though the Census Bureau stopped including the state estimates in the CPS annual report, in an effort to encourage users to wait for the ACS for information at the subnational level.

THINKING OUTSIDE THE CURRENT AMERICAN COMMUNITY SURVEY CONTENT BOX

Chester Bowie (National Opinion Research Center) presented the work he has done with Jennifer Madans (National Center for Health Statistics).

Although Madans is the primary author of the presentation, she was unable to attend the workshop.

Bowie, a former director of the Census Bureau's Demographic Surveys Division, described the ACS as a national treasure, which has the potential to serve additional uses beyond the current ones. The goal of the presentation was to envision possibilities for the ACS, setting aside currently existing limitations, such as Title 13 restrictions and consideration of the existing procedures for determining the content of the questionnaire.

Currently the Office of Management and Budget and the Census Bureau cochair an interagency committee that reviews and updates the justification for each of the questions on the ACS. The committee includes over 30 federal agencies, with OMB having the authority to make the final decision about the questions.

The interagency committee also evaluates the need for new questions, which are then tested as part of a very thorough content test scheduled to occur every five years. The content test typically involves cognitive interviews, a large-scale split-ballot field test, and follow-up interviews in some cases. The agencies also have input into the evaluation criteria during the content test.

Due to the design of the ACS, which relies on estimates over a five-year period to produce small-area data, once a question is added to the ACS, it has to remain on the survey for at least five years in order to be useful for small-area estimates. Naturally, agencies and other data users would like all of their favorite questions to stay on the ACS indefinitely. Continuity is also important from the perspective of trend data. Bowie referred back to the disability questions discussed by Schwab as an example of revisions that resulted in a break in the trend data available on this topic. The changes introduced in 2008 have affected the estimates about the populations with and without a disability.

In other words, the vision of the ACS as a resource of substantive data useful to a variety of agencies across the federal statistical system represents some practical challenges. An alternative goal for the survey would be to serve as a sampling frame for other surveys in the system. This would involve limiting the small-area data produced on the basis of the ACS to possibly only a core set of demographic variables, focusing instead on collecting data that would primarily be useful for building sampling frames. This idea builds on the examples presented by Keith Rust (see Chapter 3).

The more widespread use of the ACS as a sampling frame would still involve a difficult process of prioritizing the different agencies' needs. Even the large ACS sample may not be large enough to accommodate multiple frames for use for follow-up studies for rare populations. These populations would need to be included not only when they are the focus of a primary study but also when they are of interest as a source of sample. This would also reduce the usefulness of the ACS as a direct source of data that meet analytic needs, which could

create the appearance that the survey is less useful overall. Possible funding implications of this type of change would therefore also need to be considered.

Another alternative for the future of the ACS would be to serve as a platform for collecting data on a set of key national indicators in several areas. This would involve revising the current questions to be more in line with the core needs on various topics. For example, although there are some health-related questions currently on the ACS, these are not the most crucial set of health questions. Again, prioritizing the different agencies' needs and defining the core or key indicators would involve significant challenges and possibly the need to include a larger number of questions in some areas to meet these needs. However, Bowie argued that having each of a handful of agencies adding a small number of questions would be a realistic option that would lead to key data for small areas on a number of important topics. A variation on this approach would be to develop a core set of questions and include the expanded topic areas as modules that are on the survey for five-year periods at a time.

Bowie emphasized that the strengths of the ACS as a large-scale survey capable of providing data for the smallest geographic areas also means that prioritizing the agencies' needs and making decisions about its use will require difficult choices. He also acknowledged the possibility that the current design and scope are the most effective approach for the survey, but he encouraged participants to consider the alternatives and envision possibilities for how the different pieces in the system could fit together more efficiently.

COMPETING FEDERAL STATISTICS AND THE ROLE OF THE OFFICE OF MANAGEMENT AND BUDGET: IS THERE A NEED FOR OFFICIAL MEASURES?

Drawing on his experience, Hermann Habermann (Committee on National Statistics) discussed the role of OMB in the federal statistical system and shared his thoughts on the concept of official statistics. He remarked that OMB is a powerful institution, with authority over budgets and surveys, yet it has not assumed an active role in many years in some of the areas discussed, including greater integration in the federal statistical system and shaping official statistics. Indeed, he observed, OMB's primary role is to protect the Office of the President—in other words, to prevent bad things from happening. Activities related to this take significant time, leaving few resources for building coalitions among agencies. There is also inertia to consider and the narrowly construed "stovepipe" nature of many organizations, including OMB. Statistical policy initiatives have to cut across many agencies, each with its own separate budget.

Although it is important for OMB to provide leadership, Habermann argued that the individual agencies also have an important role to play, without which change would be very limited. He reminded participants of some remarkable initiatives that have progressed despite the challenges, including the

Confidential Information Protection and Statistical Efficiency Act (CIPSEA), the Statistical Community of Practice and Engagement (SCOPE), and work on administrative records. Nevertheless, OMB and the agencies need to be looking at the bigger picture as well as further into the future.

Habermann referred to Jelke Bethlehem's presentation about the use of population registers in the Netherlands. Although the American public tends to be less open to the concept of a national register than many other countries, the truth is that similar databases already exist, particularly at the Census Bureau. Habermann said that making better use of the information that already exists could perhaps be considered under a label such as "improving the use of existing products for statistical purposes," rather than "registers." He also said that a panel convened by the National Research Council would be in an ideal position to examine how this would work, what it would take to implement it, and how it would change the way surveys are designed. This would be one way of approaching the task of developing a new model for federal surveys systematically.

As an introduction to the topic of official statistics, Habermann recalled the 2000 census and the Census Bureau's concerns about the quality of some of the data. There was a debate whether the data were of adequate quality to be released, and in the end they were all released because the Census Bureau's mandate and obligation are to publish the data collected using taxpayer funds, provided that pledges of confidentiality are not violated. The debate made clear how difficult it is for an agency to be in a position in which it has to consider what is good enough, let alone official.

Habermann also mentioned some advantages to having the definition of a concept developed through a process outside the agency that collects the data. For example, even though the Census Bureau publishes poverty data, it does not define what poverty is—that is defined by society. This underscores the importance of considering the roles of OMB, statistical agencies, and others not only in the development of what should constitute official data, but also in defining the concepts that are measured.

In terms of competing measures of the same concept, Habermann made a case for transparency about the methodologies employed. Although he used to assume that data users want discrepant measures to be reconciled, he has learned that they often just want to understand the reasons for the discrepancies and are comfortable using competing estimates as long as the methodologies are clearly explained. This is part of the reason why competing estimates still exist on so many key topics, such as poverty estimates.

Habermann also talked about alternative data sources that often compete with official federal statistics. In several topic areas, many—if not most—of the data do not come from federal statistical agencies. For example, some of the data on environmental topics come from state, local, and nongovernmental offices. Gender statistics are often based on the work of nongovernmental orga-

nizations. Private companies are also producing their own data on a variety of topics; Google's consumer price index is a good example of this. The inclination may be to dismiss these data as less accurate than federal statistics, but it may be more productive to acknowledge these trends. Federal statistical agencies could help data users, the media, and the public better understand the data available from these additional sources. Some agencies have in fact already gone beyond this type of role, engaging in discussions with Google about making the best use of available data. Exploring possibilities for combining federal statistics with data from other sources is another topic that would benefit from the insights of a National Research Council panel.

Habermann ended by reminding participants that the goal is to move the discussion on the future of federal surveys forward. Many of the speakers before him described the challenges related to the current system, and budget pressures in the future will possibly increase these challenges further. This means that the future is likely to be different; the question is what role OMB and the statistical agencies want to play in shaping that future.

DISCUSSION

Alan Zaslavsky tied the sessions on small-area estimation and survey content together by saying that a possible criterion for deciding what should be on the ACS is to include questions that are predictive of other things for which small-area estimates are needed. What is available from administrative records and other sources could also be considered, making the ability to fill in gaps another criterion. There is no need to spend a lot of resources on collecting data that are already available.

He added that there are good reasons for Schwab's argument for measuring concepts in a uniform way if the goal is to merge the data, but for modeling purposes what is needed is the ability to link the data, knowing what the correlations are. It is not even necessary to do this every year, just once in a while to refresh information in the small-area framework.

Barbara O'Hare followed up by proposing that a question about telephone access (landline, cell, both, or neither) could be added to the ACS. Information about telephone status could then become a link to other surveys that are conducted by phone. This is especially important given budget pressures and the uncertain future of random digit dialing surveys.

Following up on a comment by Fay, Deborah Griffin (Census Bureau) clarified that the reason the Census Bureau moved away from referring to the ACS as a replacement to the census long form is because the ACS is, in fact, different, and the Census Bureau would like to be able to convey that better to data users. This does not mean that the ACS cannot be used for the same types of analysis that the long form was used for. The ACS can do much more

than what the long form could, and that is also becoming evident from this workshop.

Today the primary purpose of the ACS is to produce small-area data for the content that was inherited from the long form and that the Census Bureau promised to produce, Griffin said. The release of the first five-year estimates will represent the point at which the Census Bureau has met that goal. However, along the way, the Census Bureau has had to publish other data to keep people interested, so there were one-year and three-year estimates, and then people started talking about other things as well, such as using the data as input for SAIPE or using the survey as a test vehicle for the 2020 census or as a sampling frame. In Griffin's view, the Census Bureau staff who are currently working on the small-area estimates, who see that as the survey's mandate, are not going to want to turn the ACS into simply a sampling frame, just because it would make a good sampling frame, even if the five-year data are not going to be perfect. These ideas will have to be revisited to realign priorities, she said, and also to make sure that the ACS does not try to do so many things that it can no longer do anything well.

Daniel Kasprzyk (National Opinion Research Center) agreed with Bowie that the ACS is a national treasure, observing that the owner of that treasure has the duty to plan ahead for the ACS in ways that benefit the entire federal statistical community. Beginning the planning process cannot wait, in his view, because it will take 10 or 15 years to implement changes that are planned today.

8

Discussion and Next Steps

THE NEED FOR CHANGE

The issues and challenges facing federal data collections and the sustainability of the current system were revisited by several participants at the end of the workshop. Robert Groves said that the increasing costs of data collections combined with the possibility of declining budgets are bringing the federal statistical system to the "edge of chaos," where a small decline in a statistical agency's budget could threaten the existence of entire surveys. He argued that agencies should work together to develop contingency plans for situations in which a survey may have to be dropped, thinking about whether the statistical system collectively would still be able to produce some of the necessary data after a cut of this type. Robert Kominski voiced a similar concern, saying that federal statistical agencies tend to make decisions in a methodical and organized way, based on information available about the past. However, changes in the environment can happen, and sometimes these changes are quite large.

Graham Kalton went further, suggesting that the system is characterized by a tendency to maintain the status quo and fear of the possible adverse consequences of change. He was not sure that questioning the sustainability of the federal statistical system was warranted, but he agreed that the current surveys are not in line with many of the current needs described, especially increasing demand for data at smaller geographic areas and disaggregated for smaller subgroups to inform more focused policy-making decisions. The growth in this area has been a trend for many years, and it is time to discuss ways of addressing these needs.

Katharine Abraham agreed that the increased need for richer information

is evident from the discussions at the workshop. She emphasized that a global evaluation of the current state and the future of federal household surveys will involve making some difficult choices and setting priorities.

Kalton argued that approaching the task incrementally is quite appropriate. Groves said that, although he agrees, he would like to see a vision crystallize in the near future. Parts of a vision have seemed to emerge during the workshop and nailing that down soon would make incremental steps toward a specific vision possible. Andrew White also urged participants to spell out the intended goals and line up initiatives with their expected outcomes, especially in light of the magnitude of the projects discussed.

INTEGRATION OF SURVEY CONTENT

Abraham summarized one of the main themes of the workshop as the importance of survey content integration. One aspect of this is the use of common definitions for the concepts measured—to the extent that this is appropriate—because comparability enables researchers to make better use of the information available. Kalton said that the discussion of the development of standardized disability measures was a good example of the benefits, especially when the questions are set up so that additional measures can be added to expand the definition of a concept. The main set of questions provides a valuable benchmark for comparison across surveys.

Abraham argued that making headway in the area of integration of content would require agencies working together from the planning stages of a survey and collaborating during redesign efforts to determine crucial content. The burden cannot be placed entirely on the Office of Management and Budget (OMB). Cynthia Clark recalled her experience working on the United Nations Global Strategy to Improve Agricultural and Rural Statistics, which brought together organizations to identify the core data items that needed to be produced.

Trivellore Raghunathan compared federal statistical agencies to academic departments, in which researchers are focused on their particular disciplines. His own work illustrates that bringing together interdisciplinary teams to address these types of issues works well. This was echoed in Groves's comments that people have to stop talking to just themselves and begin a dialogue with others whom they do not usually think about when they design data collections.

Hal Stern raised the question of whether, given the costs of data collections, there is information currently collected by federal statistical agencies that goes beyond what is mandated or widely used. As an "outsider" (an academic), he said he can afford to raise difficult questions, but his question tied in with Abraham's point about addressing priorities and determining collectively which measures are crucial.

Edward Sondik (National Center for Health Statistics) also sees as valu-

able setting core standards and benchmarks for what represents critical data in a field. In the area of health, there is an explosion of information, including data collections funded by the National Institutes of Health, and many of these data collections do not go through OMB. Private companies are also producing more and more data. Sondik said that this is not necessarily good or bad, but the increase in the volume of information from an increasing variety of sources will require federal statistical agencies to step up and provide an assessment—a "consumer's report"—on the quality of these data. This is perhaps an important future role for the federal system, he said.

Kominski reminded participants that the decentralized nature of the statistical system is one of its virtues. For example, the high school dropout rate published by the Census Bureau differs from the one published by the Department of Education. This reflects differences in terms of what to measure and how to measure it, and it is not necessarily a problem, but something to consider when assessing the challenges involved in getting different agencies to coordinate their measures. He added that it is nevertheless important to ensure that coordination happens in a systematic way.

Making a similar argument as Sondik, he observed that this is particularly true in light of increasing volumes of data produced outside the federal statistical system that are receiving substantial attention, in part because they can be made available much faster than federal data. An example of this is the Google consumer price index, which is based on the tracking of online price data. Although the value and potential of these types of data are not clear, there is little doubt that researchers should at least be paying attention to these alternative approaches and that the role and usefulness of "official statistics" should be evaluated in this context as well.

Groves warned that the timeliness of data releases is a particularly big concern, because federal statistical agencies are out of sync with competing sources of information. For example, the quality of an alternative price index may be really poor, but if it is available in real time, then that may be a compelling argument for some uses. Abraham responded that a lot of the economic data are released very quickly: for example, the unemployment rate is published on the first Friday after the month to which the estimates apply, and that is quite good. Groves agreed that timeliness is relatively good in terms of the economic data released, even though he questioned why the unemployment data cannot be published weekly. However, he emphasized that in other areas the lack of timeliness is a significant problem—for example, in many cases the data released are two years old. The question becomes whether defensible estimates could be produced at a higher frequency, even if this requires more resources.

Reflecting on the topic of official statistics, Kominski argued that there are relatively few statistics that are declared official. Some are used as if they were official only because there are no alternatives available. However, having more

data on similar concepts typically leads to having to confront the question of which measures are official.

SMALL-AREA ESTIMATION

Some of the discussion revolved around the need for small-area data and modeling techniques used to produce estimates when direct estimation is not possible. Kalton clarified that the challenges in this area are usually a combination of a small-area and a small-domain problem. If the population of interest itself is small, as in the case of 5-17-year-olds in the Small Area Income and Poverty Estimates (SAIPE) Program, then the sample size of this population in a small area will also be very small. In addition, the estimate itself is often a very small proportion. These factors have consequences for modeling. He added that it is important to not lose sight of the quality of the auxiliary data used, because that is more important than the model. For example, there are distortions introduced if the data are not collected the same way in all areas, as is the case with the information about free and reduced price lunches.

Concerns were raised related to data users' willingness to embrace model-based estimates in the same way they embrace direct estimates. Kominski said that the procedures involved in SAIPE seem a little bit like "voodoo economics" to many, but focusing on educating users would go a long way toward ensuring that these types of estimates are better received. Labeling the estimates as experimental or research series would also be useful, according to Groves, who said that people need some relief from the thinking that everything published by the federal statistical system is official, because that stifles innovation. Abraham agreed, saying that when statistical agencies have gone out on a limb in the past and produced what amounts to experimental series, yet explained what they were doing clearly, the user base followed.

Another concern was the lack of statisticians with the skills required to implement advanced modeling techniques. Groves said that there is a community of people around the country who have these skills, as long as agencies are willing to look outside their existing staff and form alliances.

INTEGRATION OF SAMPLING FRAMES

Another possible direction for integration discussed at the workshop is coordination among the statistical agencies in the area of sampling frames. Clark argued that the time is right to consider the idea of a common sampling frame, and the Census Bureau's Master Address File (MAF) represents a starting point to consider. Although sharing information from the MAF outside the Census Bureau is subject to confidentiality restrictions, it is important to consider whether some parts of it are not subject to these restrictions and could be made available to other agencies under some kind of agreement. One source

of input to the MAF is the U.S. Postal Service's Delivery Sequence File (DSF); perhaps the Census Bureau could add information to it and make that product available to others.

Kalton recalled the Canadian example of the address register that is continuously updated, in part through their labor force survey. What if the United States were to bring together all of its surveys to improve an overall address frame that everyone in the statistical community could benefit from, possibly even beyond the federal statistical agencies?

Groves said that thinking about the continuous updating of the address frame does not have to be limited to the updating of addresses. Instead, it could be conceptualized as a collection of observable auxiliary data about the addresses, and various organizations could contribute information to it. Kalton added that if some of the data come from sources other than government agencies, the limitations could be different. For example, faster delivery times could be possible, and the confidentiality restrictions may also be less stringent.

THE ROLE OF THE AMERICAN COMMUNITY SURVEY

The discussions of both integration of content and sampling frames circled back to the American Community Survey (ACS) on a number of occasions. Clark said that the most important function of the ACS is to provide estimates for small areas and that it is in fact the only good source of direct estimates for small geographies. Nevertheless, other promising uses mentioned at the workshop could certainly be discussed further.

Abraham summarized the discussion about one possible use of the ACS as a more integrated household survey, with a set of rotated modules. This could increase efficiencies and lead to data that serve a broader array of analytic purposes. Clark talked about the possibility of using the ACS to help other agencies test and develop new modules. However, there are some obvious challenges emphasized by Abraham, including the burden placed on ACS respondents, the survey's inability to collect information that is comparable in depth to topic-specific surveys, and practical barriers that were brought up by the ACS team.

The possibility of using the ACS as a sampling frame for other surveys was also discussed. Clark said that this model works well in the National Agricultural Statistics Service; the Census of Agriculture accommodates screening questions for other surveys. This approach has enabled them to meet emerging needs, such as measuring bioenergy and organic production. However, she mentioned that the ACS itself in its current form has some weaknesses when it comes to rural populations, and it would not be a suitable screener for a study focused on rural America.

Kalton would have especially liked further discussion about the idea of the ACS providing sample on a rolling basis. Currently, one year's worth of ACS data has to be processed before the National Science Foundation can receive

sample for the National Survey of College Graduates (NSCG), for example. He acknowledged that providing sample on a rolling basis would involve additional data management tasks, but he thought that it was an idea worth discussing. He would have also liked more discussion of the issue of misclassification in the sample provided and how it affects a sampling design that involves rare populations.

Stern brought up the point that the ACS collects a lot of data that are not released for small areas, except after five years of aggregation. He wondered whether some of the data available could at least be used for modeling purposes, even if they are not released.

Kominski said that the ACS appeared to emerge as the silver bullet from many of the discussions, and this is perhaps not surprising given that it is a massive data system and most people have not even fully considered the power of the five-year estimates, released on a yearly basis. Even with the overlap across the data contained in those releases, 10 or 15 years of these estimates will have huge potential. However, he cautioned against limiting the thinking about the future of the federal statistical system to the ACS, especially in terms of pursuing the idea of adding modules to the survey. He used the example of the CPS, which does have supplements, but the space is booked for every month for the next three years. The CPS has been routinely used for the past 40 years by researchers both inside and outside the government as the staging ground for many new ideas and problems to be measured, and the process has been fairly efficient, but it is not an elegant method and not necessarily something that should be transferred to the ACS.

Scott Boggess (Census Bureau) reminded the workshop participants of everything the ACS is already doing. He pointed out that the ACS does in about four months what the 2000 census took approximately two years to do, and it does it with fewer resources. In addition to the long-awaited five-year estimates, they have been producing one- and three-year estimates, redesigned their weighting approach to improve variances at the tract level, redesigned their data products, developed a Spanish-language questionnaire, and added Puerto Rico and group quarters to the sample. The ACS is fast and responsive, he said, but he also made the point that it takes a long time to change an entire system.

Kalton said that the many ideas that emerged during the workshop made him question whether another survey is needed to accomplish the goals discussed. After all, the ACS has to fulfill its mandated roles before doing anything else.

ADMINISTRATIVE RECORDS

Participants were encouraged by the progress reported by Rochelle Martinez from OMB in the area of administrative records use. Clark mentioned that, while she was at the Census Bureau, she and her colleagues started the

Statistical Administrative Records System (StARS) database, and it would be of great value if that could be made available to other agencies. Some obvious uses for administrative records are direct use, imputation, verification of data, and covariates in models, but there may be others and it is important to think broadly, she said.

Kalton added that administrative records can represent a source of longitudinal data, sometimes with information available before and after the time of the survey data collection. The Panel Study of Income Dynamics (PSID) and the Health and Retirement Study (HRS), for example, use Social Security data to chart income patterns over respondents' lifetimes. Jean-Louis Tambay encouraged the participants to imagine meeting five years from now and to identify current opportunities related to administrative records that will look like a real shame to have missed looking back from the future.

Regarding the use of administrative records abroad, Kalton commented that Julie Trépanier's presentation about the use of tax records in Canada was an example of a use that reduces respondent burden and is communicated to the respondent as such. The presentation by Jelke Bethlehem about the population register in the Netherlands led to a lot of debate during the workshop, and Kalton encouraged the participants to continue that dialogue, even if a register is unlikely to be implemented in the United States in a similar form. Stern made a similar argument, saying that it is difficult to imagine that there would be political will in the United States for implementing something similar to what other countries are doing with administrative records, but that does not preclude it from discussion, because registers have the potential to offer enormous cost savings.

BROADER INTEGRATION OF DATA COLLECTIONS

A lot of the discussion centered around the more ambitious notion of integration advanced by Raghunathan, who used his own work to illustrate a way of thinking about a research problem in terms of a matrix of the information necessary to address it. Missing pieces in the matrix can be filled in with data from a variety of sources and combined using the latest modeling techniques. The analogy he drew to the statistical system as a whole generated a lot of discussion.

Abraham said that the concept of the statistical system as a giant matrix with interlocking pieces was intriguing, because it perhaps presents a solution to the dilemma of not being able to obtain all the data needed from one survey, as well as to the difficulties related to combining information from surveys that have evolved independently of each other. She emphasized that implementing something similar would require a more global way of thinking about the household surveys in the federal system. Roderick Little added that what is necessary is a new way of thinking about survey design and the associ-

ated analysis that goes beyond concentrating efforts on the specific survey one happens to be working on.

In Abraham's view, an overarching model, such as the matrix idea, would provide additional incentive for a discussion about what types of estimates are appropriate for federal statistical agencies to be generating. Sondik added that the lack of resources and capacity to produce needed small-area estimates should focus attention on defining core measures and indicators.

Kalton observed that Don Dillman's discussion of mixed-mode surveys becomes especially relevant in the context of integration among surveys. Although research has explored the effects of mixed-mode data collection within a survey, less is known about the consequences of combining data from two surveys that are conducted through different modes. The discussion of the disability measures illustrated that estimates are not necessarily the same, even when the questions are the same, and this could in part be due to a mode effect.

Kalton made the point that surveys that use other surveys as a source of sampling for rare populations could make better use of the information available from the source if there was more attention paid to coordinating content as well. In other words, if the new survey was thought of as an extension of the existing survey, then the data could be combined and used for purposes beyond what is possible with the individual surveys.

Thinking about the possibilities of linking surveys can extend beyond research domains, according to Stern. He made the point that currently surveys that rely on other surveys as a source of sample tend to do so within the same domain. An example of this is the relationship between the Medical Expenditure Panel Survey (MEPS) and the National Health Interview Survey (NHIS). Other major benefits are possible in looking beyond the institutional boundaries and to other disciplines.

According to Sondik, a report on developing key national indicators for children—which recognized that to accomplish this goal required going beyond established domains—is an example that could apply in a variety of areas, including health, education, and the economic situation. This recognition could inform more of what is done and lead to a focus on the critical information needed to serve as benchmarks. For example, the NHIS could also pick up basic information related to education and housing, in addition to its current content.

Abraham said that the initiatives in the area of administrative records also fit well with this model if one thinks beyond survey integration to envision data integration, in which administrative records are contributing an important piece. She encouraged the participants to be bold in moving forward.

References

Battese, G.E., R.M. Harter, and W.A. Fuller. (1988). An error components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association, 83*, 28-36.

Bell, B. (2010). *Combining Survey, Census, and Administrative Records Data in Small Area Models.* Presentation for the Workshop on the Future of Federal Household Surveys, November 4-5, National Research Council, Washington, DC.

Bollinger, C.R., and B.T. Hirsch. (2006). Match bias from earnings imputation in the Current Population Survey: The case of imperfect matching. *Journal of Labor Economics, 24*, 483-519.

Dillman, D. (2010). *New Data Collection Modes and the Challenge of Making Them Effective.* Presentation for the Workshop on the Future of Federal Household Surveys, November 4-5, National Research Council, Washington, DC.

Dillman, D.A., J.D. Smyth, and L.M. Christian. (2009). *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method.* Hoboken, NJ: John Wiley & Sons.

Fay, R.E., and R.A. Herriot. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association, 74*, 269-277.

Ghosh, M., and J.N.K. Rao. (1994). Small area estimation: An appraisal. *Statistical Science, 9* (1), 55-93.

Gonzales, M.E., and C. Hoza. (1978). Small area estimation with application to unemployment and housing estimates. *Journal of the American Statistical Association, 73*, 7-15.

Gonzalez, M., and J. Waksberg. (1973). *Estimation of the Error of Synthetic Estimates.* Paper presented to the International Association of Survey Statisticians, Vienna, Austria.

Hansen, M.H., W.N. Hurwitz, and W.G. Madow. (1953). *Sample Survey Methods and Theory* (volumes I and II). New York: John Wiley & Sons.

Levy, P.S., and D.K. French. (1977). Synthetic estimation of state health characteristics based on the Health Interview Survey. *Vital and Health Statistics, 2*(75). Available: http://www.eric.ed.gov/PDFS/ED156715.pdf [May 2011].

Little, R. (2006). Calibrated Bayes: A Bayes/frequentist roadmap. *American Statistician, 60*(3), 213-223.

National Research Council. (1995). *Measuring Poverty: A New Approach.* Panel on Poverty and Family Assistance: Concepts, Information Needs, and Measurement Methods, C.F. Citro and R.T. Michael, editors. Committee on National Statistics. Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2000). *Small-Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology.* Panel on Estimates of Poverty for Small Geographic Areas, C.F. Citro and G. Kalton, editors. Committee on National Statistics. Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Nelson, C. (2010). *Different Measures for Different Purposes: The Cases of Income and Poverty Measures.* Presentation for the Workshop on the Future of Federal Household Surveys, November 4-5, National Research Council, Washington, DC.

Neyman, J. (1934). On the two aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society, A*(97), 558-625.

Office for National Statistics. (2010). *Integrated Household Survey User Guide*. Available: http://www.statistics.gov.uk/downloads/theme_social/IHS/ihs-guide-complete.pdf [May 2011].

Platek, R., J.N.K. Rao, C.E. Särndal, and M.P. Singh, Editors. (1987). *Small Area Statistics: An International Symposium.* New York: John Wiley & Sons.

Purcell, N.J., and L. Kish. (1980). Postcensal estimates for local areas (small domains). *International Statistical Review, 48*, 3-18.

Raghunathan, T. (2010). *Role of Statistical Models in Federal Surveys: Small Area Estimation.* Presentation for the Workshop on the Future of Federal Household Surveys, November 4-5, National Research Council, Washington, DC.

Rao, J.N.K. (2003). *Small Area Estimation.* New York: John Wiley & Sons.

Särndal, C.E., and M.A. Hidiroglou. (1989). Small domain estimation: A conditional analysis. *Journal of the American Statistical Association, 84*, 266-275.

Särndal, C.E., B. Swensson, and J. Wretman (1992). *Model-Assisted Survey Sampling.* New York: Springer-Verlag.

Schaible, W.L., editor. (1996). *Indirect Estimators in U.S. Federal Programs.* New York: Springer-Verlag.

Schenker, N., and T.E. Raghunathan. (2007). Combining information from multiple surveys to enhance estimation of measures of health. *Statistics in Medicine, 26*, 1802-1811.

Schenker, N., J.F. Gentleman, D. Rose, E. Hing, and I.M. Shimizu. (2002). Combining estimates from complementary surveys: A case study using prevalence estimates from national health surveys of households and nursing homes. *Public Health Reports, 117*, 393-407.

Schenker, N., T.E. Raghunathan, and I. Bondarenko. (2010). Improving on analyses of self-reported data in a large-scale health survey by using information from an examination-based survey. *Statistics in Medicine, 29*(5), 533-545.

Smith, P. (2009). Survey harmonisation in official household surveys in the United Kingdom. *International Statistical Institute Invited Paper Proceedings.* Durban, South Africa. Available: http://www.statssa.gov.za/isi2009/ScientificProgramme/IPMS/0238.pdf [April 2011].

Tambay, J.-L. (2010). *Statistics Canada's Household Survey Strategy.* Presentation for the Workshop on the Future of Federal Household Surveys, November 4-5, National Research Council, Washington, DC.

U.S. Census Bureau. (2010). *American Community Survey Questionnaire*. Washington, DC: Author. Available: http://www.census.gov/acs/www/Downloads/questionnaires/2010/Quest10.pdf [April 2011].

Appendix

Workshop Agenda

Workshop on the Future of Federal Household Surveys

November 4-5, 2010
Washington, DC

Thursday, November 4

8:30-8:50   Welcome by the Workshop Steering Committee Chair
             Hal Stern, University of California, Irvine

8:50-10:10   Federal Household Survey System at a Crossroads
             Chair: Hal Stern, University of California, Irvine

                 The State of Federal Household Data Collections in the
                     United States
                 Katharine Abraham, Joint Program in Survey Methodology,
                     University of Maryland

                 Survey Harmonization in Official Household Surveys in the
                     United Kingdom
                 Paul Smith, UK Office for National Statistics
                 (Presenter: Cynthia Clark, National Agricultural Statistics
                 Service, U.S. Department of Agriculture)

99

10:10-10:25   Break

10:25-12:00   Federal Household Survey System at a Crossroads (continued)

        Statistics Without Surveys? About the Past, Present and Future of Data Collection in The Netherlands
Jelke Bethlehem, Statistics Netherlands

        Statistics Canada's Household Survey Strategy
Jean-Louis Tambay, Statistics Canada

        Discussant
Chester Bowie, National Opinion Research Center

12:00-1:00   Lunch to Continue Morning Discussion of the U.S. and International Models

1:00-2:45   Sampling Frames
Chair: Graham Kalton, Westat

        Using Large Surveys to Assist in Frame Development for Smaller Surveys
James Lepkowski, University of Michigan

        The Role of the American Community Survey in Sampling Rare Populations
Keith Rust, Westat

        Sampling Frames for Federal Household Surveys: A Vision for the Future
Frederick Scheuren, National Opinion Research Center

2:45-3:00   Break

3:00-4:40   Methods: Collection of Household Data
Chair: Katharine Abraham, Joint Program in Survey Methodology, University of Maryland

        New Data Collection Modes and the Challenge of Making Them Effective
Don Dillman, Washington State University

> Integrating Administrative Records into the Federal
> Statistical System 2.0.
> Rochelle Martinez, Office of Management and Budget
>
> Role of Administrative Records in Household Surveys:
> The Canadian Perspective
> Julie Trépanier, Statistics Canada

4:40-4:50　　Discussion of the Day's Presentations
　　　　　　　Alan Zaslavsky, Harvard Medical School

4:50-5:30　　Floor Discussion

Friday, November 5

8:30-10:15　　Methods: Small-Area Estimation
　　　　　　　Chair: Alan Zaslavsky, Harvard Medical School

> Finding the Boundaries: When Do Direct Survey Estimates
> Meet Small-Area Needs?
> Robert Fay, Westat
>
> Combining Survey, Census, and Administrative Records
> Data in Small Area Models
> William Bell, Census Bureau
>
> Role of Statistical Models in Federal Surveys: Small-Area
> Estimation
> T.E. Raghunathan, University of Michigan

10:15-10:30　　Break

10:30-12:00　　Survey Content
　　　　　　　Chair: Chester Bowie, National Opinion Research Center

> Promoting Consistency: The Case of Disability Measures
> Margo Schwab, Office of Management and Budget
>
> Different Measures for Different Purposes: The Cases of
> Income and Poverty Measures
> Charles Nelson, Census Bureau

            Thinking Outside the Current American Community Survey Content Box: What if...?
            Jennifer Madans, National Center for Health Statistics, and Chester Bowie, National Opinion Research Center

            Competing Federal Statistics and the Role of the Office of Management and Budget: Do We Need Official Measures?
            Hermann Habermann, Committee on National Statistics

12:00-1:00   Lunch to Continue Morning Discussion of Issues Related to Survey Content

1:00-1:30   Discussion of the Workshop and Next Steps
            Hal Stern, University of California, Irvine
            Katharine Abraham, Joint Program in Survey Methodology, University of Maryland
            Cynthia Clark, National Agricultural Statistics Service
            Graham Kalton, Westat

1:30-2:30   Floor discussion

2:30   Adjourn

COMMITTEE ON NATIONAL STATISTICS

The Committee on National Statistics (CNSTAT) was established in 1972 at the National Academies to improve the statistical methods and information on which public policy decisions are based. The committee carries out studies, workshops, and other activities to foster better measures and fuller understanding of the economy, the environment, public health, crime, education, immigration, poverty, welfare, and other public policy issues. It also evaluates ongoing statistical programs and tracks the statistical policy and coordinating activities of the federal government, serving a unique role at the intersection of statistics and public policy. The committee's work is supported by a consortium of federal agencies through a National Science Foundation grant.