

Discussion of Session 20

**Challenges and Opportunities in Major Survey Redesigns:
Experiences from the SIPP, NHIS, and SDR**

Patrick J. Cantwell¹

U.S. Census Bureau, 4600 Silver Hill Rd., Washington, DC 20233
patrick.j.cantwell@census.gov

In this session, we heard three excellent papers describing the redesigns in three different surveys: the National Health Interview Survey (NHIS), the Survey of Doctorate Recipients (SDR), and the Survey of Income and Program Participation (SIPP). My approach to discussing this session is to elicit common themes across the papers, and to bring in other surveys and programs that have experienced similar challenges, comparing their issues and solutions to the three we've heard. I apologize in advance that many of these examples refer to surveys conducted by the Census Bureau, where I have spent all of my Federal career. I'll consider four fundamental questions:

- Why should one redesign a survey or program?
- What should one change about the survey?
- What impediments and challenges do we face?
- Specific to these three surveys, what are some of their successes and next steps?

When addressing these questions and making decisions, it is critical to *balance* the total cost of the proposed redesign in dollars, additional time spent by staff, disruption to production processes, etc.; against the potential improvements in data products, efficiency of operations, respondent burden, the ability to integrate with other surveys or data sources, and other advances.

1. Why Redesign?

Of many potential reasons, one can focus on several of the most important ones for redesigning a survey or data-collection program. The most common reason may be to provide new or improved data products. We saw that a key interest of the SDR was to provide more detailed information on the employment status and history of doctorates and on their fine (detailed) field of degree. Other goals may be to release data earlier, or to harmonize the survey content with other data sources, as with the NHIS.

Just as common is the need to reduce the cost of the program, perhaps by reducing the size of the sample, or more generally making other changes in the design to improve efficiency. This need was a consideration for redesigning the SIPP, as well as the American Community Survey (ACS), a large household survey that provides information for states and smaller geographic areas every year. In fact, although the U.S. census of population and housing is not done on a sample basis, the Census Bureau's current mandate is to

¹ The views expressed are those of the discussant and not necessarily those of the U.S. Census Bureau.

conduct the 2020 Census at a cost no greater than that for the 2010 Census (U.S. Census Bureau 2016).

Many survey programs--including the SIPP, NHIS, and ACS--have tried to increase their response rates or reduce respondent burden, two objectives that are often closely related. The SIPP and NHIS have each shortened the questionnaire. In recent years, the ACS program conducted an extensive review, examining each question asked, and reviewing the reason and mandate for its inclusion in the survey (U.S. Census Bureau 2014). To lower the number of questions asked, the NHIS will soon include some content on its instrument in a matrix response pattern. The ACS has considered this option as well, but has not made any decision in this area.

Redesigning a survey or program provides an opportunity to incorporate newly available technologies, resources (such as administrative records), or statistical methods. As many countries conduct their census every five or ten years, the extended time between censuses allows for major changes in these areas of innovation, more so than for programs that release data every month, quarter, or year. In the United States, changes in the census planned for 2020 include the use of

- satellite imagery to correct errors in the address list from the office,
- the Internet to facilitate and increase self-response,
- administrative records to help cut the cost of visiting households that don't respond to the census, and
- an automated control and routing system to make these visits more efficient.

Of the many other reasons to redesign a survey, problems with the data produced may be a consideration. For many years, in the Monthly Retail and Wholesale Trade Surveys, the Census Bureau employed a rotating panel design in which all but the largest firms would be contacted every three months on a rotating basis. The firms would be asked to provide their sales (and, for some, their inventories) for the month that had just ended (the current month), and for the prior month. The data collected would then be combined using composite estimation, and a preliminary estimate released. A month later, when additional responses for that same data month had been collected from the next panel, a final composite estimate would be released. In the 1990s, the Bureau noted a series of cyclical revisions from preliminary to final estimate in these surveys. After extensive analysis of the response data, it was determined that the firms' responses for the current month tended to be biased downwards relative to those from the prior month. With the rotating panel design and composite estimator in place, this response bias led to inaccurate estimates and cyclical revisions (Cantwell and Caldwell 1998). To address this pattern, the Census Bureau proposed and implemented a new design when the next sample was selected.

2. What to Change?

The reason for redesigning a survey should generally be the major factor behind what gets changed. It is then not surprising that the content of the questionnaire or the resultant data products are often targeted. We saw this with the SDR, with its desire to release greater detail on employment and fine field of degree, and also with the NHIS.

To better evaluate the 2010 Census, the Census Bureau expanded the data products released based on the Census Coverage Measurement Survey (CCM). In prior decades, the focus had been on estimates of net under- or overcoverage. In 2012, based on data collected in the CCM, the Census Bureau produced estimates of the components of census coverage, such as omissions, and correct and erroneous census enumerations. At the national level, the estimate of net overcoverage was essentially 0 (0.01%; standard error, 0.14%). However, the components told the more complete story. After removing responses for which insufficient information was collected, the estimate of omissions was about 10.0 million people (std. error, 440,000). This was balanced by an estimated 10.0 million erroneous enumerations (std. error, 199,000), most of them due to duplication (U.S. Census Bureau 2012). Without these components of census coverage, first published following the 2010 Census, the estimate of net undercount might obscure the true outcome.

Over the past 40 years, the mode of response has evolved in many surveys, for reasons of cost, convenience, or data quality, and enhanced by technological developments. Some of the various innovations have included Computer Assisted Telephone Interviewing, Computer Assisted Person Interviewing, Internet response, and Computer Audio Recorded Interviewing. More recently, the SDR and other surveys have investigated the use of adaptive design and mode switching with some success. With or without changes in mode and technology, one may see changes to the questions on the instrument. As was mentioned above, the ACS recently undertook an extensive content review, studying each question, and considering the mandate and the use of the data.

Two of the surveys presented in this session--the SIPP and the NHIS--have or will change the pattern of response for some or all of their questions. The SIPP has changed from reporting every four months to reporting once each year. In the NHIS, some questions will no longer be asked every year, but in alternating patterns, depending on the importance of the topic; see "the Quilt."

In fact, when a survey sponsor believes it is time to make changes, it is not uncommon that a rotation design receives some consideration. The Current Population Survey (CPS) was perhaps the first major government survey to use a rotating design, but changed its pattern to the current 4-8-4 scheme in July 1953 (U.S. Census Bureau 2006). Just a few other surveys that have considered a rotation design in recent decades include the National Assessment of Educational Progress (Educational Testing Service and Department of Education), the Commercial Buildings and Residential Energy Consumption Surveys (Energy Information Administration, Department of Energy) (National Research Council 2012), and the SDR. On the contrary, the Monthly Retail and Wholesale Trade Surveys changed from a rotation design to a fixed-panel design in 1997 because of the problem with response bias mentioned above.

It is worth noting that some survey sponsors have altered the balance of resources expended in their attempt to improve the survey. For the 1997 version of the Commodity Flow Survey, sponsored by the Census Bureau and the Department of Transportation as part of the Economic Census, the sample was cut essentially in half--from almost 200,000 establishments in 1993 to just over 100,000 in 1997. The savings was put towards a number of activities whose focus was improving the quality of the data collected. This was the same motive for cutting the sample in the 2010 CCM from over 300,000 housing units to about 170,000. The lower cost of the sample supported expenditures toward additional training for field interviewers and increased applications of quality control (U.S. Census Bureau 2009).

3. Impediments and Challenges in Redesign

Although there is no end to potential obstacles faced when redesigning a survey, several were mentioned in this session and encountered in other programs. Perhaps the most obvious impediment is the budget required to start and complete the project. Planning change must incorporate sufficient funding, as well as time for staff, the required research, and making the appropriate decisions. Finally, money and time are necessary for implementing changes to the systems, training, and data dissemination.

Before even making a decision to redesign a survey, to predict the required funding and levels or staff, one must first develop reliable estimates of the *return on investment*. What will be the cost of the effort? How does one estimate the price of the disruption introduced by the changes? How does one measure the value of the improvements? It's hard enough to predict the true (but unknown) cost in dollars; how does one measure these other less tangible aspects of a redesign? But this must be done to weigh and balance the costs versus the benefits.

Another concern involves the integrity of a data series over time. For surveys conducted monthly, quarterly, or annually, the estimated *change* in value or percent from one period to the next is often of greater interest than the estimate of level. For example, the unemployment rate is probably the key statistic coming out of the CPS. When changes are made to the survey, it's important that any change in the unemployment rate reflects the conditions in the labor market, and not the changes to the survey. In fact, when the Bureau of Labor Statistics incorporated changes to the CPS questionnaire in 1994, it also conducted--for purposes of analysis--a separate survey using the old questionnaire. The intent was to collect data and compute the labor force estimates (with sampling error) that would have occurred had no changes been made to the questionnaire.

When redesigning, required changes to automated systems can obstruct progress. The cost of developing new systems or updating old ones usually ends up much higher than the original projection. Further, such costs often don't include all sources, such as the time spent by subject matter experts, analysts, or statisticians to explain how the operations work or what is described in requirements or specifications. Beyond that, the budgetary estimates often ignore the unmeasurable cost of disruption to the process or the working environment.

The list of challenges can be extensive. Resistance to change can be a real challenge to management interested in redesigning a program. Developing effective metrics to assess the survey or operation before and after a redesign can be difficult. Concepts such as the quality of the data produced, the coverage of a survey, or the actual time for response, are subjective, and thus open to a variety of proposed measures. In fact, a lack of consensus on the goals or even simple definitions can create a road block on the path to improvement.

As an example, consider the ACS. Faced with a number of letters from unhappy survey participants, the ACS survey managers planned and held a two-day workshop on reducing respondent burden in March, 2016 (National Academies of Sciences, Engineering, and Medicine 2016). They invited to the workshop experts in survey methodology and experienced users of ACS data to present their views, experience, research results, and proposals. However, it became clear during the opening sessions that the invitees could not agree on how one defines respondent burden. Is it the number of questions on the survey instrument? If the survey contains 40 questions, does one count questions 31a, 31b, and

31c as one question or three? Perhaps the average length of the interview in minutes is more appropriate. Or should it be the average length per household person, or something in between? Is respondent burden a function of the difficulty answering the questions? Do you need records, such as a pay stub (electronic or on paper) to determine income? Does the sensitivity of the question or concerns about privacy affect the perception of respondent burden? Does annoyance play a role? Should the average or maximum number of contact attempts be a factor? Is it more annoying when the interviewer calls up to 12 times, rather than stopping after three attempts? Should respondent burden include the level of disruption to the household contacted? Is it less disruptive or burdensome to be contacted by mail or the internet? One can toss the piece of mail onto the pile and complete the questionnaire later when one has time. The same could be said for e-mail. But it can be more irritating to sit down for dinner and hear the telephone ring or the knock on the door. In summary, how does one define--and then measure--respondent burden, or many other multi-faceted concepts?

4. Successes, Moving On

Finally, let us return to the three survey programs presented in this session, and review the successful implementation of their redesigns. We heard how the Census Bureau made major design changes to the Survey of Income and Program Participation. For the first time since the survey's inception decades ago, the reporting pattern was changed from every four months to once a year. After many years of testing and study, the survey has incorporated an event history calendar to help with recall. Administrative records are being used in several areas. The redesign benefited as the Census Bureau involved its data users in the process and considered their needs. Most important, the SIPP management kept its focus on the basics of the survey. As they considered changes, they kept their commitment to making SIPP the primary source of data on sources of income, participation in Federal and other government programs, and flows in and out of these programs or statuses. The remodeled SIPP guarantees this commitment into the future.

The Survey of Doctorate Recipients expanded its sample and changed the selection process for the 2015 survey. But additional changes are due in the near future. A large cohort of new sample cases was introduced in 2015. With so many new cases--more than three quarters of the sample--a lack of contact information was a bigger problem than in prior survey years. This problem is exacerbated by the number of older doctorates for which a current e-mail address is not available. But the National Science Foundation (NSF) has adjusted its data collection procedures to accommodate these issues. Further, the new procedures will yield improved data on the employment of doctorate recipients, and better information on their fine field of degree--two primary objectives of the survey redesign. They engaged their stakeholders and listened to their suggestions for improvement. As the SDR incorporates the additional changes, the NSF is considering the users' need for better longitudinal data.

For the first time in 20 years, the National Health Interview Survey has successfully changed its content, with an eye toward the future and the changing state of data collection in household surveys. The National Center for Health Statistics (NCHS) realized that the average time to complete the survey had increased, and response rates had decreased. They introduced a survey instrument with rotating content (the Quilt), allowing them to shorten the questionnaire and reduce respondent burden. Changes were not made unilaterally, but after consultation with the survey's stakeholders, and discussions of their data needs. The NCHS made a difficult choice in that some data will not be available every year. But their

decision process successfully considered the balance of input factors: the data content to be released, the resulting response burden, and the budget.

Once again, I congratulate the authors on their excellent research, implementation of the new designs, and presentations today.

References

Cantwell, P. and Caldwell, C. (1998). "Examining the Revisions in Monthly Retail and Wholesale Trade Surveys Under a Rotating Panel Design." *Journal of Official Statistics*, 14, No. 1, 47-59.

<http://www.jos.nu/Articles/abstract.asp?article=14147>

National Academies of Sciences, Engineering, and Medicine (2016). *Reducing Response Burden in the American Community Survey: Proceedings of a Workshop*. Washington, DC: The National Academies Press. doi: 10.17226/23639.

<https://www.nap.edu/catalog/23639/reducing-response-burden-in-the-american-community-survey-proceedings-of>

National Research Council (2012). *Effective Tracking of Building Energy Use: Improving the Commercial Buildings and Residential Energy Consumption Surveys*. Panel on Redesigning the Commercial Buildings and Residential Energy Consumption Surveys of the Energy Information Administration. W.F. Eddy and K. Marton, Eds. Washington, D.C.: the National Academies Press.

U.S. Census Bureau (2006). *Current Population Survey: Design and Methodology*. Technical Report 66. October, 2006.

<https://www.census.gov/prod/2006pubs/tp-66.pdf>

U.S. Census Bureau (2009). "Recommendation to Reduce Nonsampling Error in the 2010 Census Coverage Measurement Program." DSSD 2010 Census Coverage Measurement Memorandum Series #A-43. Prepared by David C. Whitford.

U.S. Census Bureau (2012). "2010 Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States." DSSD 2010 Census Coverage Measurement Memorandum Series #G-01. Prepared by Thomas Mule.

http://webdev.ssd.census.gov/coverage_measurement/pdfs/g01.pdf

U.S. Census Bureau (2014). "Release of the American Community Survey (ACS) Fiscal Year 2014 Content Review Results Final Report and Supporting Documentation." ACS Information Memoranda Series Memorandum No. 2014-02. From James B. Treat to The Record.

<https://www.census.gov/programs-surveys/acs/operations-and-administration/2014-content-review.html>

U.S. Census Bureau (2016). *2020 Census Operational Plan: A New Design for the 21st Century*. Released October 2016. Version 2.0. Washington, D.C.

<https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-oper-plan2.pdf>