

Modernizing a Major Federal Government Survey: A Review of the Redesign of the Current Population Survey Health Insurance Questions

Joanne Pascale¹

Measurement error can be very difficult to assess and reduce. While great strides have been made in the field of survey methods research in recent years, many ongoing federal surveys were initiated decades ago, before testing methods were fully developed. However, the longer a survey is in use, the more established the time series becomes, and any change to a questionnaire risks a break in that time series. This article documents how a major federal survey – the health insurance module of the Current Population Survey (CPS) – was redesigned over the course of 15 years through a systematic series of small, iterative tests, both qualitative and quantitative. This overview summarizes those tests and results, and illustrates how particular questionnaire design features were identified as problematic, and how improvements were developed and evaluated. While the particular topic is health insurance, the general approach (a coordinated series of small tests), along with the specific tests and methods employed, are not uniquely applicable to health insurance. Furthermore, the particular questionnaire design features of the CPS health module that were found to be most problematic are used in many other major surveys on a range of topic areas.

Key words: Health reform; questionnaire redesign; health insurance; CPS.

1. Introduction

Measurement error – the difference between the “true value” of a concept being measured and the survey estimate that represents that concept – can be very difficult to assess and reduce. While there have been great strides in the field of survey methods research in recent years, many ongoing federal surveys were initiated decades ago before testing methods were fully developed. As such, the degree and nature of measurement error associated with these surveys is often unknown. However, the longer a survey is in use, the more established the time series becomes, and any change to a questionnaire risks a break in that time series. Thus, a comprehensive survey redesign is generally approached with caution for a number of reasons. First, demonstrating that any changes actually do reduce measurement error (and do not inadvertently introduce other problems) can be elusive. Second, even if a statistic is imperfect (e.g., a point estimate is biased), if the cause of the imperfection does not interact with time, then the trend line of that point estimate can still be valid and informative. Third, the loss of the time trend to the data-user community due

¹ Center for Survey Measurement, US Census Bureau, 4600 Silver Hill Road, Suitland, MD 20233, U.S.A. Email: joanne.pascale@census.gov

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

to a break in series is sometimes untenable. Finally, research to bridge the break in series is costly and sometimes data do not exist for the task. On balance, continuing with the status quo can be the best course of action in many cases.

This article documents a middle-ground approach in which a key component of a major federal survey – the health insurance module of the Current Population Survey Annual Social and Economic Supplement (CPS ASEC, aka CPS) – was redesigned over the course of 15 years through a series of small, iterative tests, both qualitative and quantitative. The CPS is one of several surveys (federal, state, and private) that measure health insurance coverage, but the CPS is the most widely cited and used source of estimates on health coverage in the United States (Blewett and Davern 2006). This is partly due to its trend line on coverage going back to the 1980s, and because the sample size is large enough to make state-level estimates (Farley-Short 2001). It also offers rich and detailed auxiliary data on income and employment, making it a particularly valuable data source for researchers investigating connections between economic well-being, health coverage and health status. The CPS, however, has also been the subject of widespread criticism because its estimate of the uninsured tracks higher than that of other major surveys (Bhandari 2004; Congressional Budget Office 2003). For example, the estimate of those uninsured throughout calendar year 2012 was 15.4 percent in the CPS and 11.1 percent in the National Health Interview Survey (NHIS). The NHIS also produces an estimate of those uninsured at a single point in time, and in 2012, it was 14.7 percent, which happens to be close to the CPS 2012 calendar-year estimate of 15.4 percent (State Health Access Data Center 2013). Indeed, other studies have also found that the CPS calendar-year estimate was very similar to point-in-time estimates from other surveys (Rosenbach and Lewis 1998; Congressional Budget Office 2003), leading to widespread speculation over what the CPS estimate really represented – calendar year or point-in-time coverage or something in between (Lewis et al. 1998; Swartz 1986).

Due to the divergent estimates across surveys, and the persistent criticism of the CPS, in 1999 the Census Bureau began a comprehensive research program to examine and reduce measurement error associated with health insurance estimates from the CPS, focusing on the role of the questionnaire. Numerous small-scale studies – both qualitative and quantitative – were fielded and analyzed, and results were fed into subsequent small-scale tests in an iterative fashion. The research approach was to identify features of the questionnaire that were potential candidates for contributing to measurement error, explore and modify those features, and test against the status quo to assess empirical evidence for improvements due to the changes. After a decade of this testing, a fundamentally redesigned questionnaire was crafted. A formal pretest of the redesign was conducted in 2009 and the basic approach was found to be sound (Pascale 2009a). Minor refinements were made and a large-scale split-ballot experiment was conducted in March 2010 with promising results (Boudreaux et al. 2013). By chance, the very day the 2010 test was launched (March 23, 2010), the Patient Protection and Affordable Care Act (ACA) was passed. In response to that, the draft redesign was then adapted with questions specific to health reform and tested in 2011–2012 with residents of Massachusetts, given its passage of ACA-like state-level legislation in 2006. That testing proved successful (Pascale et al. 2013), and the adaptations were integrated into the redesigned questionnaire for a follow-up large-scale field test in 2013. Results were favorable to the redesign (Pascale et al. 2015),

and it was officially implemented in the March 2014 CPS ASEC data collection. See Figure 1 for a side-by-side display of the basic question structure for the old and new CPS.

This article documents the 15-year history of the CPS ASEC health insurance module redesign. The tests and results are summarized and synthesized with other relevant survey methods literature to illustrate how particular questionnaire design features were identified as problematic, and how improvements were developed and evaluated. In terms of evaluation criteria, it is important to note here that the CPS and most surveys define health

Old CPS	New CPS
<ol style="list-style-type: none"> 1. In 2013 was anyone in hh covered by job plan? <ul style="list-style-type: none"> • Yes →2 • No →6 2. Who in hh were policyholders? →3 3. Who else in hh was covered? →4 4. Did plan cover anyone outside hh? →5 5. Did emp/union pay all/part/none of premium →6 6. In 2013 was anyone in hh covered by direct plan? <ul style="list-style-type: none"> • Yes →7 • No →10 7. Who in hh were policyholders? →8 8. Who else in hh was covered? →9 9. Did plan cover anyone outside hh? →10 10. In 2013 was anyone in hh covered by outside hh? <ul style="list-style-type: none"> • Yes →Who? →11 • No →11 11. In 2013 was anyone in hh covered by Medicare? <ul style="list-style-type: none"> • Yes →Who? →12 • No →12 12. In 2013 was anyone in hh covered by Medicaid/state-specific program? <ul style="list-style-type: none"> • Yes →Who? →13 • No →14 13. How many months in 2013 was NAME covered? →14 14. In 2013 was anyone in hh covered by CHIP? <ul style="list-style-type: none"> • Yes →Who? →15 • No →15 15. In 2013 was anyone in hh covered by military? <ul style="list-style-type: none"> • Yes →Who? →Plan type →16 • No →16 16. In 2013 was anyone in hh covered by other/state-specific program? <ul style="list-style-type: none"> • Yes →Who? →Plan type →17 • No →END 17. I have recorded NAME(s) not covered; correct? <ul style="list-style-type: none"> • Yes →END • No →Who was covered? →Plan type →END 	<p>CK1: If disabled or age=65+ →1; else →2</p> <ol style="list-style-type: none"> 1. Are you covered by Medicare? <ul style="list-style-type: none"> • Yes →11 • No →2 2. Are you NOW covered by any type of health plan? <ul style="list-style-type: none"> • Yes →3 • No →Qs on Medicaid, CHIP, state-specific program names, verification of uninsured →15 3. Is it provided thru a job, govt, or other way? <ul style="list-style-type: none"> • Job →6 • Government →4 • Other way →7 4. Is that plan related to a JOB with the government? <ul style="list-style-type: none"> • Yes →6 • No →5 5. Is that Medicare, Medicaid/CHIP, military, other? <ul style="list-style-type: none"> • Medicaid/CHIP/other/DK →9 • Military →[type of military plan] →10 • Medicare →11 6. Is the plan related to military service in any way? [if yes, type of military plan] →10 7. How is it provided – parent/spouse, direct, other? <ul style="list-style-type: none"> • Parent/spouse/direct →10 • Other →8 8. Is it thru former emp, union, group, assn, school? <ul style="list-style-type: none"> • Former emp/union/group/assn/school →10 • Other →9 9. What do you call the program? [pick list of state-specific program names] →11 10. Who is the policyholder? [hh roster; outside hh] [If Q7=parent/spouse: thru their job or direct?] →11 11. Did coverage start before January 1, 2013? <ul style="list-style-type: none"> • Yes →12 • No →Qs on start month/year →12 12. Has coverage been continuous since then? [if no →Qs on months of spells] →13 13. Is anyone else in hh covered on same plan? [if yes, who?] →14 14. Were they covered same months? <ul style="list-style-type: none"> • Yes →15 • No →What months were they covered? →15 15. Any [other] coverage Jan 2013 till now? <ul style="list-style-type: none"> • Yes →loop thru series again, starting with 3 • No →CK2 for next person on roster <p>CK2: if any coverage already reported for person →15; else → CK1: If no more on roster →END</p>

Fig. 1. Old versus New CPS Questionnaire Structure (mimicked for March 2014 administration)

insurance by asking about coverage through a range of different sources or types of comprehensive coverage, both public (e.g., Medicaid) and private (e.g., employer-sponsored insurance). This is sometimes referred to as the “laundry-list” approach. Respondents are asked about each type of coverage and those without any of the listed types of coverage are then tabulated as uninsured (Lewis et al. 1998). While the particulars of this list vary across surveys, most exclude specialty plans (e.g., dental, vision, accident plans) and definitional differences have been found to have negligible empirical effects on the uninsured estimate (Farley-Short 2001). In the course of the CPS redesign development, some of the tests included a validation component for reports of particular types of coverage, but there simply is no single, comprehensive, accurate source of data on those with and without coverage in the U.S. that could serve as a gold standard for validating the uninsured estimate. Given the literature cited above suggesting that the CPS is particularly prone to underreporting of coverage (Congressional Budget Office 2003; Bhandari 2004; Lewis et al. 1998; Swartz 1986), a default “more is better” model was used to assess reduction in measurement error in the uninsured estimate.

The particular subject of this research is health insurance, but the general approach (a coordinated series of small, iterative tests), along with the specific tests and methods employed, is not uniquely applicable to health insurance. Furthermore, the particular questionnaire design features of the CPS health module that were found to be most problematic are used in many other major surveys on a range of topic areas. Thus, the article concludes with a discussion of the broader implications of this research for other topic areas and for survey redesign efforts in general. For more details, see Pascale (2015).

2. Methodological Variations in Data Sources

For many major social indicators – for example, poverty, disability, and labor force participation – there are multiple data sources, each with their own purpose and constraints in terms of data collection. That is, the purpose and budget of the survey determine key design features such as content, sample size, mode, timing, and frequency of data collection. Variation in these design features can lead to variation in the estimates, leaving researchers to weigh out the strengths and weaknesses of the measures, as well as other factors for analysis such as sample size and auxiliary variables on the dataset. As a result, often the data source that researchers come to rely on for a particular analysis is not purpose built for the specific concept under study.

In the case of health insurance, several major federal surveys came to include questions on health coverage gradually, as the health care system in the United States changed and as data needs developed. Taking them chronologically, the National Health Interview Survey (NHIS), sponsored by the National Center for Health Statistics, has been fielded since 1957 for monitoring the health of the U.S. population. In 1959, questions on private coverage were added, and in the early 1970s questions were added on large public programs that had been introduced in the 1960s – Medicaid (primarily for low-income individuals), and Medicare (primarily for those 65+) (Blumberg 2014). Starting in 1981, the CPS ASEC began including questions on certain types of health insurance and by 1987 the questions collected data on Medicaid, employer-sponsored insurance (ESI) and other

noncash benefits in order to inform the poverty measure (US Census Bureau 2015a). In 1983, the Census Bureau launched the Survey of Income and Program Participation (SIPP), the main purpose of which is to measure the dynamics of economic and social well-being over time. As such, the focus is on measuring household income and benefits from a comprehensive range of sources, including health insurance (US Census Bureau 2015b). In 1997, the Agency for Healthcare Research and Quality launched the National Medical Care and Expenditure Survey (NMCES), which was replaced by the Medical Expenditure Panel Survey (MEPS) in 1996. The MEPS collects data on the usage, cost and financing of health services, as well as health insurance (Agency for Healthcare Research and Quality 2015). Finally, the American Community Survey (ACS), designed to replace the decennial census long form, began in 2005 and health insurance questions were added in 2008 (US Census Bureau 2015b).

While the breadth of data sources on health insurance has its advantages, one downside is that the methods and the estimates they produce vary across surveys and it is not clear which estimates are most accurate (Office of the Assistant Secretary for Planning and Evaluation 2005). Research going back to the 1980s suggests that much of the variation in estimates across these surveys is rooted in subtle differences in the questionnaires (Swartz 1986). This notion prompted the current line of inquiry investigating the association between questionnaire design features and measurement error. The overarching finding from this inquiry is that within the questionnaire, three fundamental design features are particularly important in terms of their potential for driving the estimates. One is the “reference period” – the time period specified in the survey question. Some surveys ask about current coverage status, while others ask about coverage over a certain time span. The reference period is then intertwined with the definition of the uninsured. For example, the CPS is administered in March and asks respondents if they had coverage “at any time” during the previous calendar year. The uninsured are then defined as those uninsured throughout the entire calendar year. The NHIS and ACS, on the other hand, both ask about current coverage and define the uninsured as those without coverage at a particular point in time (i.e., the day of the interview). The MEPS and the SIPP, both longitudinal surveys that follow respondents for a number of years, use a reference period somewhere in between. Both surveys ask about monthly coverage during the reference period and the uninsured can then be defined in several different ways – uninsured in any given month, throughout the calendar year, throughout an entire three-year panel, or any number of months in between.

Another difference across surveys is the specificity of questions about household members. For most plan types, the CPS, NHIS, and MEPS use a “household-level” approach: “. . . was anyone in this household covered by [plan type X]?” If yes, follow-up questions are asked to determine which household members have the coverage. The SIPP and ACS, on the other hand, employ a “person-level” approach and ask about each household member by name (e.g., “Does [name] have [plan type X]?”). The SIPP, furthermore, attempts self-response for all household members age 15 and older, while in the ACS, CPS, MEPS, and NHIS a single household respondent is asked questions on behalf of all household members.

A third major design feature that varies is the specificity of questions concerning the type of health coverage. The CPS, MEPS, and ACS ask a series of yes/no questions, each

on a particular type of health coverage. The NHIS takes a different approach and asks a global question about any coverage and, if yes, a single follow-up question is asked to determine the particular type of coverage.

3. Problematic Design Features and Prototype Improvements

3.1. Reference Period

For at least two decades, researchers have been trying to understand the source of variation in health-coverage estimates. For example, Swartz's seminal article in 1986 examined the sampling framework, weighting procedures, adjustments for nonresponse and attrition, and questionnaire design across four different surveys (CPS, SIPP, NHIS, and the NMCES). Swartz concluded that the largest contributor to variation in the estimates was differences in the questionnaire, particularly differences in the reference period (Swartz 1986). Research comparing the calendar-year and current reference periods corroborated Swartz's findings (Rosenbach and Lewis 1998; Pascale 2001a), and prompted qualitative research to understand more about measurement error associated with the reference period. This research found that some respondents simply do not hear or do not focus on the reference period stated in the question, and instead report on their current situation (Pascale 2008). A similar investigation of the twelve-month reference period in the context of a CPS supplement on Food Stamps receipt also found a lack of attentiveness to the reference period (Hess and Singer 1995). Related literature showed that respondents tend to underreport receipt of benefits from the more distant past (Lynch 2006; National Research Council 2006; Resnick et al. 2004; Ringel and Klerman 2005), which is consistent with findings from the survey methods literature more generally (Schaeffer and Presser 2003). Quantitative studies on Medicaid corroborated the finding on underreporting of coverage in the more distant past, and also found that accuracy of past coverage improved if respondents were currently covered (Pascale et al. 2009; Research Project to Understand the Medicaid Undercount 2008). In sum, it appears there were several factors indicating that the CPS phrase "at any time during [past year]" often failed to prompt recall of events going back to the beginning of the calendar year.

A subtle but potentially important point about the reference period has to do with "lag time." This is the length of time between the interview date and the time period of coverage asked about in the survey question. In surveys with a current reference period, obviously, there is no lag time. However, in the CPS there is roughly a three-month lag time which, in light of research findings that some respondents focus on their current situation, could compound the problems with recalling and reporting past coverage. Respondents are never anchored in their current day-of-interview status but are given the task of thinking back over 15 months (from January of the previous year until March of the current year), focusing on the first twelve of those months while "subtracting out" the most recent three months. Furthermore, they are asked about coverage "at any time" during those twelve months which, technically speaking, includes coverage for as little as one day. Thus the relatively long duration of the reference period, the three-month lag time, and the fact that respondents are not asked about their current situation may all be working against the CPS.

To address these issues, an exploration of ways of asking about both current and past calendar-year coverage within the same set of questions was undertaken. The rationale was to accommodate respondents' tendencies to focus on their current status (even if instructed otherwise) and then leverage that current status as an anchor to ask about retrospective coverage. Memory and recall literature suggested that providing multiple time frames could enhance the reporting accuracy of past events (Crespi and Swineheart 1982; Sudman et al. 1984; Loftus et al. 1990; Blair and Ganesh 1991; Martin et al. 2002; Pascale 2009b). Putting the health-coverage findings together with the memory literature resulted in a question series with two time frames: the date of the interview, and the beginning of the reference period (January 1 of the previous calendar year). This approach both anchors the respondent in the present day and frames the full 15-month time span of interest.

Very generally, the series begins by asking about current coverage, and then asking whether that coverage started before January 1 of the previous year. If it did, respondents are asked if the coverage has been continuous since January 1. If so, it is inferred that the coverage was held all 15 months. If the coverage began after January 1, respondents are asked in what month the current coverage started. These respondents, along with those whose coverage was not continuous, are asked about specific months within the reference period for which they were not yet reported to have coverage. This is a key departure from the much more general line of questioning asking "At any time during [past year] . . ." that was found to be problematic in the recall literature.

3.2. Household- Versus Person-Level Questions

While reference-period issues have tended to dominate the literature on health insurance measurement error, as a design feature, reference period alone does not explain all the variation observed in the estimate of the uninsured. For example, both the SIPP and the MEPS employed reference periods shorter than the calendar year, yet the difference in their uninsured-throughout-the-year estimates for 2002 was quite striking: 8.1 percent in the SIPP and 12.9 percent in the MEPS (Davern 2009). Key differences between these surveys are the household- versus person-level approach (used in the MEPS versus SIPP, respectively), and the fact that the SIPP is a self-response survey while the MEPS asks one household member to answer questions about all other household members. The combination of the person-level approach and the self-response could account for SIPP's higher reporting of coverage relative to the MEPS.

Most surveys do not have the luxury of self-response, but the difference in the MEPS and SIPP uninsured estimates suggests there could be some benefit to "naming names" – that is, asking the household respondent about household members by name, versus a more general "anyone in the household." There is, indeed, some evidence that a failure to name each household member individually risks the respondent failing to report coverage for some members, particularly in larger or complex households (Blumberg et al. 2004; Hess et al. 2001). On the other hand, administering the entire series for each household member individually risks respondent fatigue and associated underreporting (Blumberg et al. 2004; Pascale 2001b). So while there are pros and cons to both the household- and person-level approaches, it is not entirely clear how the overall estimates are affected across all plan types and across households of various sizes and complexity.

The first experiment on the CPS redesign integrated both the household/person-level design and reference-period features in order to isolate the effects of each. This 1999 experiment employed a two-by-two design in which respondents were randomly assigned to one of four treatments: (A) calendar year/household-level (this is the status quo CPS ASEC); (B) calendar year/person-level; (C) current coverage/household-level; and (D) current coverage/person-level. In the absence of measurement error, there should be no difference in estimates within the calendar-year designs or within the current-year designs. That is, reporting should be the same whether questions are asked at the household- or person-level. However, asking about coverage “at any time” during the past calendar year should, theoretically, result in higher reporting than asking about current coverage, since the former allows for coverage on ANY of 365 days of the year and the latter only includes one – the day of the interview. Results from the experiment were telling. Within the current reference period versions, the difference between person- and household-level designs was nonsignificant – that is, there was no evidence of measurement error ascribable to the household/person-level design when asking about current coverage status. However, when asking about past calendar-year coverage, the uninsured rate in the household-level design was 5.1 percentage points higher than in the person-level designs ($p < 0.01$). This suggested that respondents could report on all household members, whether or not they were provided with the individual names, equally well when they were only asked about current coverage, but when asked about coverage during the past calendar year they reported more coverage when provided with individual household members’ names (Pascale 2001a).

Within the person-level versions, the difference in reporting was observed in the expected direction. The insured rate was higher in the calendar-year design than in the current design by 3.4 percentage points ($p < 0.05$). However, within the household-level designs, the insured rate was identical – at 12.0 percent – for both calendar-year and current designs. This suggested the reference-period wording – “at any time during [previous calendar year]” – was effective at eliciting reports of past coverage when respondents were asked to think about only one person at a time, but not when they were asked to think about “anyone in this household.” In short, there appeared to be some kind of cognitive overtaxing going on within the household/calendar-year design in the CPS.

To address this, a hybrid household-person-level approach was developed. Each household member is asked about by name, but whenever a specific plan or plan type is reported, a follow-up question is asked to determine whether anyone else in the household is also covered by that same plan or plan type. That information is stored and harnessed so that when subsequent household members are asked about, if they were already reported to have coverage, the question series is significantly abbreviated and only asks about any additional coverage. This allows the questions to reference each household member by name (rather than the more general “anyone in the household”), but it avoids the tedium of repeating the entire series for each household member. It also avoids redundant reporting of health plans in the vast majority of cases where multiple household members share the same type of coverage. Finally, this change, in combination with asking about current and past coverage in separate questions, greatly reduces the risk of overtaxing the respondent by loading a single question with too many demands (thinking about ALL household members, across ALL 15 months).

3.3. Questions on Individual Coverage Types

With regard to the specificity of questions on coverage type (the laundry-list approach noted above), a number of reporting problems has been identified in the literature (Beatty and Schechter 1998; Loomis 2000; Roman et al. 2002; Pascale 2009c; Pascale 2008; Schaeffer and Presser 2003; Willson 2005). First, the list itself is often not mutually exclusive. For example, employer-sponsored insurance (ESI) and military coverage can overlap, and coverage from someone outside the household can overlap with ESI and/or directly purchased coverage. Second, respondents can have difficulty figuring out which category best fits their coverage. For example, self-employed individuals who get coverage through a trade association are sometimes torn between the ESI and directly purchased category. In addition, dependents on a spouse or parent's job-based plan are sometimes reluctant to report their coverage as ESI because the coverage is not through *their* job but that of the spouse or parent. Third, individual questions on plan type are often too detailed and complex for respondents to grasp with confidence, or they fail to tap into the respondent's understanding of the coverage. This proved particularly problematic when respondents were answering for other household members, about whom they had only limited knowledge of personal circumstances like health coverage. For example, in many cases, respondents knew another household member was covered, and that it was a government-sponsored plan, but they were unclear on the distinction between Medicaid and Medicare (Loomis 2000; Roman et al. 2002; Willson 2005). All these factors led respondents to misreport one plan type as another, report the same plan twice, or fail to report the plan altogether (Loomis 2000; Pascale 2008).

One of the most compelling examples of this type of misreporting is demonstrated in two related studies. Loomis (2000) conducted cognitive testing of the CPS series but used two alternative versions – one with the status-quo method in which Medicare was asked about prior to Medicaid and one that reversed just the order of questions on those two plan types within the series. She found that respondents confused the two programs, which sometimes meant they reported the same coverage twice (e.g., Medicaid enrollees reported their coverage at both the Medicare and the subsequent Medicaid question). Loomis concluded it “. . . seems quite possible for Medicaid recipients to simply respond ‘yes’ to the first question that sounds familiar to them.” (Loomis 2000, 16).

Following on from these findings, a split-ballot field test on sequencing effects of Medicare and Medicaid was conducted in 2003 in which the standard CPS series of questions was asked, but in half the sample the Medicare question preceded the Medicaid item and in the other half the sequence of these two plan types was reversed. Results suggested there was false-positive reporting of Medicare when that plan type was asked first. Among low-income households (in which individuals are more likely to be covered by Medicaid), when Medicare was asked first, the Medicare estimate was 24.8 percent, but when Medicaid was asked first the Medicare estimate was 18.6 percent. The Medicaid estimate was unaffected by question sequencing (Pascale 2004). A later validation study of people known through Blue Cross/Blue Shield records to be enrolled in Medicaid found that “a fair number of Medicaid enrollees – when asked both the Medicaid and Medicare question – answered yes to both” (Davern et al. 2008). Though this one example is suggestive of a certain pattern of misreporting, overall it is difficult to gauge the magnitude

and direction of misreporting stemming from the laundry-list approach. However, there is ample evidence from this and numerous other studies of the potential for underreporting, double reporting and misreporting of plan type (Pascale 2009c).

The redesign addresses the problems with the laundry-list approach by starting with a relatively simple question about whether the respondent has coverage or not, and following up with questions that go from general to specific to identify plan type. The objective was to make each individual question easier for respondents to understand and answer, and to tap into the level of knowledge they did have. For respondents who report some kind of coverage, a follow-up question first determines the general source of coverage – through a job, the government or state, or some other way – and tailored questions from each response category obtain the necessary detail. For job-based plans, subsequent questions identify policyholders and dependents, and determine if the coverage is related to military service in any way. For government plans a follow-up question asks about the type of government plan, presenting both Medicare and Medicaid in the same list of response categories, so that respondents can assess which plan type is closest to their understanding of the coverage they have. For respondents who choose “other” as their general source of coverage, follow-up questions ask about plans obtained in the next most common ways (other than employment and government) – through direct purchase and, to accommodate dependents on someone else’s private plan, through a parent or spouse. In the end, the same level of detail (in fact, more detail in some cases) is captured, but in a way that is more respondent friendly. The questions go from very general to very specific in terms of plan type, and the series accommodates respondents’ sometimes-limited knowledge of other household members’ health-coverage situation.

4. Testing of the Redesign

4.1. Cognitive Testing and Pretest (2008–2009)

The initial prototype of the redesigned questionnaire addressed the reference period, household-level design and the laundry-list approach. Informal testing of this experimental draft was conducted with an unpaid convenience sample in order to correct any fatal flaws before conducting cognitive testing with paid respondents. Cognitive testing was carried out in the spring of 2008 with 36 household respondents (Pascale 2009a), followed by a larger pretest in March 2009 with 54 household respondents. Minor refinements were made along the way and iteratively tested, and there was no evidence of problems with these changes (Pascale 2009b). Only one substantive design issue was identified. Technically speaking the aim of the CPS ASEC is to capture coverage during the previous calendar year, and one major goal of the redesign was to address the distinction between current and past-year coverage. Thus initial redesign questions captured only past calendar year and day-of-interview coverage. However, pretest results suggested it was, perhaps ironically, less burdensome to collect data across the entire 15-month time span with wording such as “And was [NAME] also covered from January, 2008 up until now?” than it was to ask about current coverage and then only the previous calendar year. The post-pretest questionnaire, then, was modified slightly for the next field

test to ask about continuous coverage – from January 1 of the previous calendar year up to the day of the interview (that is, including the first few months of the current year).

4.2. *Survey of Health Insurance and Program Participation (2010)*

The 2010 split-ballot field test of the old vs new CPS, called the Survey of Health Insurance and Program Participation or SHIPP, was carried out in the spring of 2010 by the Census Bureau's telephone-interviewing staff. The sample was drawn from two sources – an RDD frame and Medicare enrollment files. Medicaid records were sought but unavailable, so Medicare records were used given the scarcity of validation studies in health-measurement research. Individuals under 65 and those recently enrolled were oversampled under the assumption that these groups were more vulnerable to misreporting – and hence more could be learned from them – than those 65 and over and/or enrolled for longer periods of time. Data were collected on 8,507 individuals, split about 60/40 across the RDD/Medicare samples. Response rates (based on the AAPOR RR4 definition) were 47.6 percent and 61.4 percent for the RDD and Medicare samples, respectively.

Results showed that there were no statistically significant differences in calendar-year estimates of the uninsured, or estimates of coverage by plan type, between the old and new CPS (Boudreaux et al. 2013). However, the direction of differences favored the new CPS, and coverage and sample bias could have been major contributors to the lack of statistical significance in the differences. For cost reasons, the SHIPP survey was entirely landline-telephone-based and did not include a cell-phone-only or face-to-face component, introducing an overall bias to both questionnaire panels. In the latter half of 2010 almost 30 percent of households in the U.S. were cell-phone only, and individuals living in those households were more likely to be young adults (in particular, those aged 25–29), living in or near poverty, living with unrelated adult roommates, Hispanic, male, and uninsured (Blumberg and Luke 2011). This makes it difficult to draw conclusions from the unweighted SHIPP data with regard to any absolute estimates, but relative comparisons should be valid since assignment to treatment was randomized. However, these relative comparisons could be compromised if the methods operate differently among subgroups that were particularly lacking in the overall sample. For example, one of the goals of the redesign was to encourage more accurate reporting of past coverage – particularly relatively short spells of coverage – by prompting respondents with the specific months for which no coverage had yet been reported. If respondents with short, intermittent spells of coverage were underrepresented in the SHIPP sample overall, then this potential advantage of the redesign would be statistically imperceptible.

The SHIPP data were also evaluated for face validity of person/month/plan-level data. Statistical analysis was limited by the small sample size, coverage bias, and the low prevalence of change in coverage status during the 15- to 17-month reference period. However, the patterns exhibited were informative. Among the nonelderly sample who were administered the new CPS ($n = 2,882$), the vast majority (80.3 percent) were insured throughout the entire reference period, 11.2 percent were uninsured throughout, and another 2.5 percent had coverage in January 2009 but lost it and never regained it by the end of the reference period. The remaining 6.1 percent ($n = 176$) began a spell of insurance during the reference period, and five of these individuals began two spells of

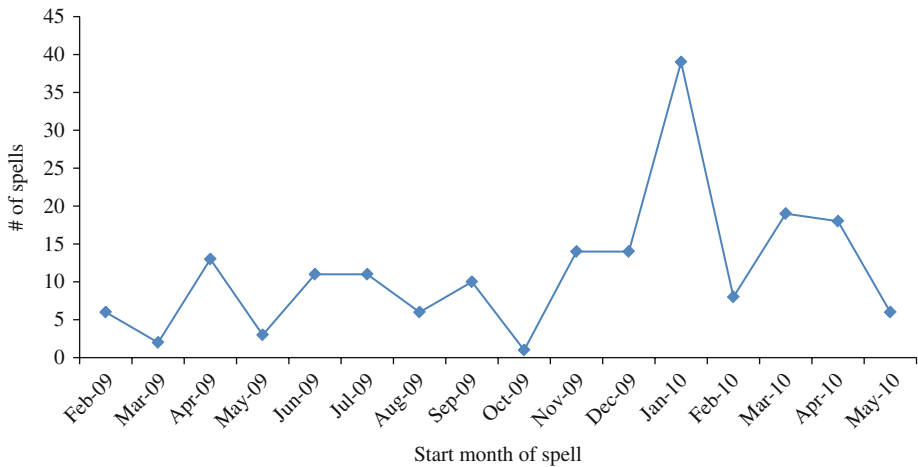


Fig. 2. SHIPP 2010 Start Month of Insured Spells ($n = 181$ spells). Source: 2010 Survey of Health Insurance and Program Participation

coverage during that time. Figure 2 plots the start month of these 181 spells. On average, just over eleven spells began in any given month. The range went from a low of one spell (beginning in October 2009) up to a high of 39 spells (beginning in January 2010) (Pascale 2011). The uptick in January 2010 may reflect actual behavior due to open enrollment, New Year’s resolutions and/or other events tied to the start of the calendar year. This uptick notwithstanding, there does not appear to be evidence of seam bias or other obvious systematic biases in terms of reported start date of spells across the 15-month reference period.

The Medicare portion of the SHIPP survey data was matched back to the Medicare records from which that sample was drawn originally, and indicators of enrollment from the two data sources were compared. Results were favorable to the redesign in terms of both underreporting and overreporting, but for the latter results were particularly robust. The false-positive (overreporting) error rate in the old CPS was 6.9 percent and in the new CPS the error rate was 2.0% — a 4.9 percentage-point difference ($p = 0.01$) (Resnick 2013).

It was hoped that the new CPS would garner higher reports of past coverage than the old CPS. However, the new CPS did not lose ground compared to the status-quo design, and coverage bias could explain the lack of differences in the estimates. Furthermore, the person/month/plan-level data had face validity, and the Medicare match-back study favored the redesign. All these factors warranted a follow-up test.

4.3. CPS ASEC Content Test (2013)

In March 2013 the Census Bureau carried out the CPS ASEC 2013 Content Test to evaluate the CPS redesign in a production environment with a larger sample ($n = 29,629$) more representative of the CPS. The study was a comparison of estimates from the old and new CPS questionnaire. Estimates for the old CPS were derived by subsetting the CPS production sample to those interviewed by phone in March 2013. The new CPS was administered in parallel with the production CPS, by phone, to a sample of households that had already completed the final rotation of the CPS (this is known as “retired” sample).

Weights, multivariate modeling, and a separate analysis of a subset of the retired sample from the production side were all used to assess the effects of the nature of differences across samples. Results showed that the odds of coverage being reported for the past calendar year were higher under the new CPS than the old, and that within the new CPS, calendar-year estimates of coverage were higher than and distinct from point-in-time estimates. There were few statistically significant differences in coverage across demographic subgroups. See [Pascale et al. \(2015\)](#) for a thorough description of the weighting, methodology and results. These results demonstrated that the redesign did represent an improvement in the estimates, and there was no evidence that subgroups were unevenly affected by the design difference.

Further methodological analysis of the same dataset was conducted in an attempt to assess the impact of the individual features of the questionnaire that were changed. This was not a straightforward process, however, because all three questionnaire design features were modified at the same time in order to evaluate the final questionnaire as a whole. Two characteristics in particular, however, were tied to some of the modified questionnaire features: household size and “social distance” – the relationship between the household respondent and the person for whom he/she was reporting (aka self/proxy). The hybrid person-household-level design in the new CPS could reduce the chances that certain individuals are forgotten, especially in larger households, because it provides names of each household member. The new CPS structure of questions on coverage type (general to specific) allows respondents to provide basic information on whether other household members are covered or not (even if their knowledge of coverage type is limited), while the old CPS laundry list asks very detailed questions on coverage type but no general yes/no question on coverage. The more social distance, the less knowledge respondents may have on the details of coverage type for other household members. This could lead to more underreporting for more socially distant individuals in the old CPS than the new. While the reference-period changes could not be completely disentangled from the other two questionnaire features, comparing estimates from single-person households at least removes the effect of the household/person-level design for this subset. In both the old and the new CPS, the questions refer to only the household respondent (“you”) and no names are used (see [Figure 3](#), upper panel). Thus, any observed differences in single-person households could be attributed to either the reference-period change, the change to the laundry-list approach, or some combination.

In [Table 1](#), two distinct variables are examined. The first is a measure of household size, where households were categorized as small (single-person), medium (two to four people) or large (five or more people). The second is a measure of social distance, which categorizes individuals as having either a self- or proxy report of their health coverage. Within proxy reports, individuals were classified based on their relationship to the “reference person,” who is usually the household respondent or their spouse. The proxies were divided into two groups: the child of the reference person, or someone other than the child of the reference person. Chi-squared tests of association were used to test the statistical significance of bivariate comparisons of coverage across test and control panels. A logistic regression model was used to test within-treatment comparisons between small, medium, and large households and between self- and proxy reports. For all comparisons, a significance threshold of 0.05 was used.

	Old CPS	New CPS
Single-person households	<ol style="list-style-type: none"> At any time in 2009 were you covered by [plan type x]? At any time in 2009 were you covered by [plan type y]? [repeat for all eight plan types] 	<ol style="list-style-type: none"> Do you NOW have any coverage? Yes → [Qs to identify plan type] → Q2 No → next section Did your coverage from [plan type x] start before or after January 1, 2009? Before → Q3 After → Q4 And has it been continuous since January 2009? Yes → CK No → Q4 In what month did that coverage start? → Qs on gaps in coverage prior to start month → CK CK: If single-person household → next section; else → Q5
Multi-person households	<ol style="list-style-type: none"> At any time during 2009 was anyone in this household covered by [plan type x]? Yes → go to Q1a No → go to Q2 1a. Who was covered? → Q2 At any time during 2009 was anyone in this household covered by [plan type y]? Yes → go to Q2a No → go to Q3 2a. Who was covered? → Q3 <p>[repeat for all eight plan types]</p>	<p>[Same routine as in single-person households, with the addition of Q5-Q8]:</p> <ol style="list-style-type: none"> Is anyone else in this household also covered by [plan type x]? Yes → Q6 No → next section Who is covered? → Q7 Were they also covered in [months first person had the coverage]? Yes → apply same months to this person as first person No → Q8 What months between January 2009 and now was this person covered by [plan type x]?

Fig. 3. Simplified Question Flow by Household Size in Old versus New CPS

Regarding household size, Table 1 shows that among single-person and large households, there was no difference between the old and new CPS insured estimate. However, in medium households, the new CPS insured rate was 2.4 percentage points higher than in the old CPS design ($p = 0.004$). Within the new CPS there was no difference in the estimates between small and medium households, but within the old CPS the insured estimate in small households was 2.7 percentage points higher than in medium households ($p = 0.021$). In both the old and new CPS, reporting of coverage was much lower in large than in small households, by almost eight percentage points. Reporting of coverage was also lower in large households compared to medium households in both treatments ($p < 0.001$ in the new CPS; $p = 0.006$ in the old CPS).

Turning to self/proxy results, Table 1 shows there was no treatment difference in the insured estimate for self-responses, but for proxy responses overall the new CPS insured rate was two percentage points higher than the old CPS ($p = 0.041$). Among proxies, there was no treatment difference in reporting of coverage for the child of the reference person, but proxy reporting of coverage for any other household member was 3.6 percentage points higher in the new versus old CPS ($p = 0.003$). Within the new CPS, there was no difference between self and proxy reporting, whether the proxy was the child of the reference person or someone else in the household. But, within the old CPS the estimate for self-reports was 2.5 percentage points higher than for proxy reports overall ($p < 0.001$). Furthermore, among proxies, the old CPS reporting was

Table 1. Weighted Old versus New CPS Insured Estimates, by Household Size and Self/Proxy Status, 2013 CSP ASEC Content Test

Household Size	New CPS Insured		Old CPS Insured		New-Old % insured (p)
	N = 243,474,924	% (SE)	N = 243,474,924	% (SE)	
Small	23,878,488	90.74 (1.0152)	27,571,316	91.24 (0.9246)	-0.50 (0.7136)
Medium	155,833,875	90.96 (0.5383)	150,507,500	88.57 (0.5784)	2.39 (0.0043)
Large	37,932,138	82.76 (1.7761)	36,233,716	83.62 (1.8462)	-0.87 (0.7532)
Self/Proxy	86,334,860	89.98 (0.5656)	88,093,242	89.52 (0.5553)	0.45 (0.5676)
Proxy	131,309,641	89.01 (0.6735)	126,219,290	87.00 (0.6535)	2.00 (0.0408)
Child	66,524,680	89.72 (0.9987)	63,078,540	89.50 (0.8721)	0.22 (0.8743)
Other	64,784,961	88.29 (0.7498)	63,140,750	84.65 (0.8696)	3.64 (0.0026)

SE are clustered on household. Unweighted n = 29,629 (new CPS sample n = 16,401; old CPS sample n = 13,228). Weighted totals do not sum to population estimates due to the sample design of the 2013 content test and drop-out rates within the ASEC portion of the questionnaire. See Pascale et al. (2015) for more details. Source: 2013 CPS-ASEC Content Test.

4.9 percentage points higher for child of the reference person than others in the household ($p < 0.001$). These results suggest that the new CPS closed the social-distance gap that existed in the old CPS.

In sum, if we assume higher reporting is more accurate reporting, the old and new CPS both do equally well in single-person households and when the respondent has to report only for him/herself and/or his/her child. Given the large gap in reporting between small and large households, one could also say both treatments do equally poorly once the household gets to be five or more people in size. However, it is unknown whether coverage levels are simply lower in larger households, or whether both treatments are missing reports of coverage for some individuals in large households. There is some evidence to support the former. For example, Hispanics as a group have higher uninsured rates than the overall population (19.9 percent versus 10.4 percent) (US Census Bureau 2014). And where the householder is of Hispanic origin the household size is larger, on average, than the national average (3.54 versus 2.59 people per household) (US Census Bureau 2012).

Where the new CPS demonstrates significant, measurable improvements is in medium-sized households and with proxy reports for individuals who are the most socially distant from the respondent. To estimate how many additional people were reported as insured under the new CPS solely due to its different impact on these two subgroups, a rough calculation is offered using the 2013 content test data. For reasons described elsewhere (Pascale et al. 2015), the 2013 test sample was weighted to a total somewhat lower than the U.S. population – about 243 million – and it is on that base population that these calculations are offered. Individuals in medium-sized households constituted 70 percent of the sample, those in the proxy/other category constituted 30 percent of the sample, and 25 percent of the total sample ($n = 60,868,731$) was in both groups. The new-old CPS difference in the insured estimate was 2.4 percentage points for individuals in medium households, and 3.6 percentage points for proxy/other individuals. To be conservative, if we allocate all of the 25-percent overlapping sample to medium households (where the new-old CPS difference is lower), the calculation would be:

Medium-sized households in the old CPS

- 169,925,637 individuals * 2.39 ppt new-old CPS difference \rightarrow 4,061,223

Proxy/Other individuals in the old CPS

- 74,590,668 individuals – 60,868,731 (the 25% overlapping sample) = 13,721,937
- 13,721,937 individuals * 3.64 ppt new-old CPS difference \rightarrow 499,479

Based on these rough calculations, a total of about four and a half million individuals reported as uninsured in the old CPS would have been reported as insured in the new CPS. Taking this a step further, if we add this four and a half million to the total reported as insured under the old CPS, the new hypothetical insured total comes up to 218,873,234, or 89.9 percent of the total old CPS population. This compares to the insured rate of 88.0 percent under the old CPS and represents a 1.9 percentage-point difference. This difference is in line with findings from the 2013 content test, which showed a difference of 1.4 percentage points in the uninsured rate between the old and new CPS. It is also in line with a comparison of the 2013 and 2014 production estimates, which employed the old and new CPS, respectively. While there were many caveats to that comparison, a 1.6 percentage-point difference was approximated to be attributable to the change in questionnaire design (Pascale et al. 2015).

4.4. Questionnaire Administration Time in the Old Versus New CPS

Timers were built into both the 2010 and 2013 tests. In 2010, the old CPS health insurance module took two minutes and 25 seconds to administer, while the new CPS took on average three minutes 56 seconds – about a minute and a half longer. The median measure showed a gap of about one minute between the old and new CPS. The 2013 results were almost identical, indicating that the new CPS took 1.5 minutes longer than the old CPS health insurance module (Bee and Cantu 2013). While the added duration in the new CPS is concerning, there are some important mitigating factors. For example, the average SHIPP interviewer had 8.39 years of field interviewing experience at the Census Bureau, and 5.41 years of experience on surveys that include questions about health insurance (the old CPS being among the most common of these). For the new CPS, interviewers only had the benefit of a two-day training and a ten-day field period. As interviewers become more familiar with the new CPS – which is a tremendous departure from the old design – the learning curve benefits should thus develop and reduce administration time. Other important considerations are the benefits of the redesign discussed below.

5. Discussion

When this research began in 1999, the objective was to identify key questionnaire design features that were associated with measurement error, develop improvements, and test their effectiveness. The 2013 test results indicated that the new CPS did reduce measurement error in estimates, across demographic subgroups, resulting in a lower estimate of the uninsured. Further analysis of the 2013 test, focusing on household size and self/proxy reporting, shed some light on the reasons for the observed improved reporting. While long-standing speculation suggested that recall error induced by the reference period was the main contributor to reporting error, results indicate that much more than recall error was at play. If recall were the only factor, then we would expect improved reporting under the new CPS even in single-person households, but we see no differences there.

What seems to have been driving lower reporting in the old CPS was not primarily recall error but a confluence of factors related to household size and composition, and cognitive difficulty of the reporting task. Consider that when asked in the old CPS, “At any time in 2012 did anyone in the household have [Plan Type X]?”, respondents who fail to report coverage are not saying “no” – that an individual household member does NOT have that coverage type – they are just not saying “yes.” Furthermore, they are failing to say “yes” to two different kinds of questions. First, in cases where the coverage type specified in the household-level question is unfamiliar or obscure to them (e.g., Medicaid and its state-specific program names, Tricare and other military plans) they may not say “yes” simply because they do not recognize the coverage type. Second, if they do answer “yes” to the household-level question, they then have to actively report individual household members who have that coverage type, and do not have the benefit of being asked about those individuals by name. In these cases, the more socially distant household members may not come to mind as readily. For example, a respondent reporting ESI coverage may be focused mostly on their own family members who are dependents on their particular plan, and less focused on other, more socially distant household members (e.g., a cousin, boarder), who may have their own ESI plans. Indeed, social distance has been found to

affect data quality in other contexts as well. For example, [Grieco and Armstrong \(2014\)](#) reported that social distance between the respondent and others in the household was likely a factor in item nonresponse to questions on year of naturalization.

Social distance, however, is only part of the reporting problem. Results indicate that when the reporting task becomes particularly challenging – calling to mind several other household members at once, some of whom are more socially distant, AND thinking back over 15 months, sometimes about obscure coverage types – the old CPS design is less equipped than the new design to prompt reporting. These results hearken back to the 1999 study on reference period and household versus person-level design. That study suggested that respondents had difficulty when asked about both the past calendar year *and* all household members at once. They had less difficulty when either one person at a time was asked about (by name) for the calendar year, or when only the current time period was asked about for “anyone in the household” (with no names). All three revised questionnaire features in the new CPS play a role in simplifying the reporting task and collectively reducing the cognitive burden on the respondent. The series asks about only one person at a time, and it starts with a simple yes/no question on current coverage status – not specific coverage type. Furthermore, the series starts with questions about not just any household member but the respondent him/herself. This allows the respondent to focus on what are likely the more difficult aspects of the questions – coverage type and months of coverage – under (arguably) the simplest circumstances first, without being burdened with *also* thinking about those same aspects for other people over a 15-month time period.

Apart from addressing measurement error, the new CPS renders all the same data as the old CPS but with much more detailed information on monthly coverage. When health insurance questions were first added to the CPS in the 1980s, the strategy was to be consistent with the unit of measurement used for income questions. Since then the landscape of health insurance has changed, which raises the question: what kind of definition of the uninsured “should” be used? That is, what degree of “uninsured” makes most sense as an analytic category? In the old CPS, the only information collected was whether individuals had coverage “at any time” during the previous calendar year. Thus, the only way to define the uninsured was those without coverage throughout the entire year because that was the only data rendered by the questions. As such, one day of coverage could separate the insured from the uninsured. This is a rather dubious distinction, since a person who was uninsured for, say, eleven out of twelve months, or who had spells of noncoverage, is likely to have a profile more like a person uninsured throughout the year than someone who was insured for the entire twelve months. Indeed, according to the Congressional Budget Office, “Policies aimed at increasing coverage are most likely to be effective if they consider the distinction between the short-term and long-term uninsured.” ([Congressional Budget Office 2003](#), viii).

The original redesign strategy of asking about current coverage, framing the 15-month time period and, in some cases, prompting respondents with month-level questions on coverage, was all in the service of improving reports of retrospective coverage. However, as a side benefit, the questionnaire renders person/month/plan-level data, from the beginning of the previous calendar year all the way through and including the interview day. Thus, the redesign enhances the value of the dataset by enabling a more thorough analysis of the dynamics of coverage over a continuous 15-month reference period.

For example, analysts can examine the number and duration of spells, and the specific months of coverage, which allows for studies on topics such as seasonality and the effects of major external events, such as a recession. The data also enable analysis of churning within plan type (e.g., people who are on and off Medicaid), transitions from one plan type to another, and the direction of those transitions (e.g., from Medicaid to ESI, or vice versa). And while the traditional CPS does allow for the reporting of multiple plan types, because it only asks about coverage “at any time” during the past year it is impossible to know the duration of coverage, and whether individuals were covered by multiple plan types at the same time or transitioned from one plan type to another. Having the data to study source of coverage and transitions will likely become more valuable in the coming years. As the uninsured rate drops, the research community focus may shift from measuring the uninsured to exploring how individuals obtain their coverage, and the dynamics of coverage – when and why they shift from one source to another.

Like all studies that aim to measure the uninsured, definitive conclusions regarding measurement error are saddled by lack of a truth source. The U.S. health care system, both pre- and postreform, is a patchwork of private and public sources of coverage and no centralized database on individuals with and without health coverage exists. However, some limited conclusions can be drawn about the data quality of the new CPS. First, studies on the old CPS indicated the calendar-year estimate of the uninsured was in line with estimates of those uninsured at a point in time. In the new CPS, the finding that the estimate of calendar-year coverage is higher than and distinct from the estimate of current coverage represents an improved metric, at least in the relative sense. Second, the new CPS appears to be less prone to measurement error in relation to household size and social distance. Third, to the extent that relevance is a measure of data quality, the improved precision and detail in terms of person/month/plan-type-level variables, and the additional ACA-specific measures in the new CPS, represents a marked improvement over the old CPS.

The next steps in this line of research include continued study of public coverage reporting. The new CPS was designed to address the chronic and persistent underreporting of public coverage, but appears not to have made any gains in that area. Analysis of the 2013 test results suggests this could be a result of multiple factors all revolving around the theme of overreporting (Pascale et al. 2015). Results from the 2010 test showed that reporting of both public and private coverage was more than double in the old CPS than it was in the new CPS (Boudreaux et al. 2013), and that Medicare overreporting was higher in the old than in the new CPS (Resnick 2013). Furthermore, although the Medicaid undercount is well documented (Klerman et al. 2005; Blumberg and Cynamon 1999; Czajka and Lewis 1999; Lewis et al. 1998), there is some evidence of Medicaid overreporting (Klerman et al. 2009; Davern et al. 2008). The latter study showed that roughly 20 percent of the self-reported Medicaid estimate was comprised of enrollees in private coverage who also reported Medicaid. Finally, the dearth of evidence on overreporting is not indicative of a lack of overreporting; rather, it is much more difficult to measure due to state-level variation and the absence of a truly comprehensive dataset of enrollees.

Another key area for more research is validation. The accuracy of the uninsured estimate is intertwined with the accuracy of reporting on individual plan types. Medicaid reporting has received substantial study and attention, in part due to the existence and

accessibility of fairly high-quality records. Yet even within the Medicaid reporting literature it is not entirely clear how misreporting of Medicaid affects estimates of other plan types, and the ultimate measure of the uninsured, at the national level. The accuracy of reporting of private plans has received less rigorous study, in part due to less accessible, more disparate sources of validation data. Finally, only two studies to date (Davern et al. 2008 and Nelson et al. 2003) have examined reporting accuracy of multiple sources of coverage – both public and private – within the same questionnaire.

To address this gap, the Census Bureau collaborated with other agencies to conduct a study in 2015 called CHIME (Comparing Health Insurance Measurement Error) that compared data from enrollment records across multiple markets – including ESI, Medicaid and the new marketplace plans – to survey reports from health insurance modules from the CPS ASEC redesign and the ACS. Analysis will include an assessment of both “absolute” reporting accuracy (the extent to which the survey data matches the record data) and “relative” reporting accuracy (comparing absolute reporting accuracy across questionnaire treatments) of particular plan types, both private and public.

6. Implications Beyond Health Insurance Measurement

These results have broad implications for both questionnaire redesign strategies and for topic areas beyond health insurance. Many redesign efforts involve a long-term plan whereby each phase of the research and the schedule is set years in advance. This often means the timeline is insufficient to fully digest the results of one study in order for the findings to be fed in to the design of the next study. More important, it can often mean that results from an early study raise questions that warrant a subsequent test of a certain nature and scope, but a different type of test was predefined in the research plan. Some degree of planning and scheduling is obviously necessary for staffing and budgeting purposes. However, if considerable flexibility is built in to the research plan, the value of each test can be enhanced with sufficient time for analysis, and each test can be adjusted to address the emerging research questions.

Results also provide guidance for questionnaire design, as the specific features found to be problematic are not unique to health insurance questions. The overarching finding is that if a reporting task is particularly demanding, it can pay to decompose complex questions into their simpler parts. Knowing what constitutes “particularly demanding” for all respondents is certainly not possible. However, results provide some guidance on how to design questions so that reporting under the more challenging conditions does not suffer relative to the simpler conditions. Numerous surveys ask retrospective questions using wording like “At any time during [time period X] did [you/anyone] receive [subject Y]?” These surveys may benefit by decomposing the task and asking first about the present (i.e., “Do you receive [subject Y]?”), which allows respondents to focus on subject Y without the compound tasks of thinking about duration and any change over time within the same question. In terms of subject matter, many topic areas lend themselves to being asked at a general level (e.g., retirement plans) or a specific level (e.g., 401(k), 403(b), IRA, Roth IRA). Results suggest asking about the general level first and then drilling down to the specifics, particularly when a single respondent is answering questions about all household members. Decomposing complex questions has been demonstrated to result in

higher data quality in contexts beyond health coverage (Loftus et al. 1990; Redline 2013; Fowler 2004). Note, however, that the evidence is somewhat more mixed when the measures in question relate to frequency of behaviors (Beatty 2010; Belli et al. 2000; Schaeffer and Presser 2003).

Finally, to reduce the risk of underreporting by providing individual names, while also reducing burden, results suggest that a hybrid person-household-level approach can be effective. This could be especially relevant when asking about a topic area that may be shared across household members, such as jointly owned assets, or household-level receipt of public benefits.

7. References

- Agency for Healthcare Research and Quality. 2015. "Medical Expenditure Panel Survey, MEPS Background." Available at: http://meps.ahrq.gov/mepsweb/communication/household_participant_back.jsp (accessed on October 5, 2015).
- Beatty, P. 2010. "Considerations Regarding the Use of Global Survey Questions." Paper presented at the Consumer Expenditures Survey Methods Workshop, Hyattsville, MD, December 8–9, 2010. Available at: http://www.bls.gov/cex/methwrkshp_pap_beatty.pdf (accessed on February 5, 2016).
- Beatty, P. and S. Schechter. 1998. "Questionnaire Evaluation and Testing in Support of the Behavioral Risk Factor Surveillance System (BRFSS), 1992–98." Office of Research and Methodology, NCHS, Working paper series, 26: 12–17.
- Bee, C.A. and A. Cantu. 2013. "Using Timer Data to Evaluate the Respondent Burden of the 2013 CPS ASEC Content Test." Proceedings of the Federal Committee on Statistical Methodology Research Conference, November 4–6, 2013, Washington, DC. Available at: http://fscm.sites.usa.gov/files/2014/07/H3_Bee_2013FCSM.pdf (accessed on February 5, 2016).
- Belli, R.F., N. Schwarz, E. Singer, and J. Talarico. 2000. "Decomposition Can Harm the Accuracy of Behavioural Frequency Reports." *Applied Cognitive Psychology* 14: 295–308.
- Bhandari, S. 2004. "People with Health Insurance: A Comparison of Estimates from Two Surveys." Survey of Income and Program Participation (SIPP) Working Paper No. 243. Washington, D.C.: U.S. Census Bureau. Available at: <https://www.census.gov/sipp/workpapr/wp243.pdf> (accessed on February 2, 2016).
- Blair, E.A. and G.K. Ganesh. 1991. "Characteristics of Interval-Based Estimates of Autobiographical Frequencies." *Applied Cognitive Psychology* 5: 237–250. Doi: <http://dx.doi.org/10.1002/acp.2350050306>.
- Blewett, L.A. and M.E. Davern. 2006. "Meeting the Need for State-Level Estimates of Health Insurance Coverage: Use of State and Federal Survey Data." *Health Services Research* 41: 946–975. Doi: <http://dx.doi.org/10.1111/j.1475-6773.2006.00543.x>.
- Blumberg, S. 2014. "National Health Interview Survey Since 1957." Federal Statistics on Health Insurance Coverage: Technical Meeting on Methods Used in Household Surveys, August 18, 2014, Washington, DC. Available at: http://www.census.gov/newsroom/releases/pdf/20140818_gw_final.pdf (accessed on February 5, 2016).

- Blumberg, S.J. and M.L. Cynamon. 1999. "Misreporting Medicaid Enrollment: Results of Three Studies Linking Telephone Surveys to State Administrative Records." Proceedings of the Seventh Conference on Health Survey Research Methods. 189–195. Available at: http://www.cdc.gov/nchs/data/hsrhc/hsrhc_7th_proceedings_1999.pdf. (accessed on February 5, 2016).
- Blumberg, S.J. and J.V. Luke. 2011. "Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, July-December 2010." National Center for Health Statistics. Available at: (<http://www.cdc.gov/nchs/nhis.htm>) (accessed on February 5, 2016).
- Blumberg, S.J., L. Osborn, J.V. Luke, L. Olson, and M.R. Frankel. 2004. Estimating the Prevalence of Uninsured Children: An Evaluation of Data from the National Survey of Children with Special Health Care Needs, 2001. Vital Health Statistics 2 (136). Hyattsville, MD: National Center for Health Statistics.
- Boudreaux, M.H., B. Fried, J. Turner, and K.T. Call. 2013. "SHADAC Analysis of the Survey of Health Insurance and Program Participation." State Health Assistance Data Center. Available at: http://www.shadac.org/files/shadac/publications/SHIPP_final_report.pdf (accessed on August 4, 2014).
- Congressional Budget Office. 2003. "How Many People Lack Health Insurance and for How Long?" A CBO Report. The Congress of the United States. Available at: <https://www.cbo.gov/publication/14426> (accessed on February 5, 2016).
- Crespi, I. and J.W. Swineheart. 1982. "Some Effects of Sequenced Questions Using Different Time Intervals on Behavioral Self-Reports: A Field Experiment." Paper presented at the American Association for Public Opinion Research. Hunt Valley, MD, May 1982.
- Czajka, J. and K. Lewis. 1999. "Using Universal Survey Data to Analyze Children's Health Insurance Coverage: An Assessment of Issues." Washington, DC: Mathematica Policy Research, Inc.
- Davern, M. 2009. "Unstable Ground: Comparing Income, Poverty & Health Insurance Estimates from Major National Surveys." Paper presented at the Academy Health Annual Research Meeting, June 29, 2009, Chicago.
- Davern, M., K.T. Call, J. Ziegenfuss, G. Davidson, T. Beebe, and L. Blewett. 2008. "Validating Health Insurance Coverage Survey Estimates: A Comparison of Self-Reported Coverage and Administrative Data Records." *Public Opinion Quarterly* 72: 241–259.
- Farley-Short, P. 2001. "Counting and Characterizing the Uninsured." University of Michigan, Ann Arbor, MI: Economic Research Initiative on the Uninsured Working Paper Series. Available at: <http://www.umich.edu/~eriu/pdf/wp2.pdf>. (accessed on August 25, 2005).
- Fowler, F.J. 2004. "The Case for More Split-Sample Experiments in Developing Survey Instruments." In *Methods for Testing and Evaluating Survey Questionnaires*, edited by S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, and E. Singer. 173–188. Hoboken, NJ: Wiley.
- Grieco, E.M. and D.M. Armstrong. 2014. "Assessing the 'Year of Naturalization' Data in the American Community Survey: Characteristics of Naturalized Foreign Born Who Report – and Don't Report – the Year They Obtained Citizenship." Paper presented at the Applied Demography Conference, April 30 – May 3, 2014, San Antonio, TX.

- Hess, J., J. Moore, J. Pascale, J. Rothgeb, and C. Keeley. 2001. "The Effects of Person-level vs. Household-level Questionnaire Design on Survey Estimates and Data Quality." *Public Opinion Quarterly* 65: 574–584.
- Hess, J. and E. Singer. 1995. "The Role of Respondent Debriefing Questions in Questionnaire Development". In Proceedings of the Section on Survey Research Methods, August 13–17, 1995. 1075–1080. Washington, DC: American Statistical Association. Available at: http://www.amstat.org/sections/srms/proceedings/papers/1995_187.pdf (accessed on February 5, 2016).
- Klerman, J.A., M. Davern, K.T. Call, V. Lynch, and J. Ringel. 2009. "Understanding the Current Population Survey's Insurance Estimates and the Medicaid 'Undercount'" *Health Affairs* web exclusive. Doi: <http://10.1377/hlthaff.28.6.w991>.
- Klerman, J.A., J.S. Ringel, and B. Roth. 2005. "Under-Reporting of Medicaid and Welfare in the Current Population Survey". RAND Working Paper WR-169-3. Santa Monica, CA: RAND.
- Lewis, K., M. Ellwood, and J. Czajka. 1998. *Counting the Uninsured: A Review of the Literature*. Washington, DC: The Urban Institute.
- Loftus, E.F., M.R. Klinger, K.D. Smith, and J. Fiedler. 1990. "A Tale of Two Questions: Benefits of Asking More Than One Question." *Public Opinion Quarterly* 54: 330–345.
- Loomis, L. 2000. "Report on Cognitive Interview Research Results for Questions on Welfare Reform Benefits and Government Health Insurance for the March 2001 Income Supplement to the CPS." Washington, DC: Center for Survey Methods Research, Statistical Research Division, U.S. Census Bureau.
- Lynch, V. 2006. "Causes of Error in Survey Reports About Who in the Household Gets Welfare." Unpublished Paper, Joint Program in Survey Methodology. College Park, MD.
- Martin, E.A., R.E. Fay, and E.A. Krejsa. 2002. "Analysis of Questionnaire Errors in Survey Measurements of Census Coverage." In Proceedings of the American Statistical Association Survey Research Methods Section, August 10–15, New York, NY. 2260–2265. Alexandria: VA American Statistical Association.
- National Research Council. 2006. *Food Insecurity and Hunger in the United States: An Assessment of the Measure. Panel to Review the U.S. Department of Agriculture's Measurement of Food Insecurity and Hunger*, edited by G.S. Wunderlich and J.L. Norwood. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Nelson D.E., E. Powell-Griner, M. Town, M.G. Kovar. 2003. A Comparison of National Estimates From the National Health Interview Survey and the Behavioral Risk Factor Surveillance System. *American Journal of Public Health* 93:1335–1341.
- Office of the Assistant Secretary for Planning and Evaluation, Health and Human Services. 2005. "Understanding Estimates of the Uninsured: Putting the Differences in Context." Updated September 5. Available at: <http://aspe.hhs.gov/health/reports/05/uninsured-understanding-ib/#estimates> (accessed on April 9, 2016).
- Pascale, J. 2001a. "Methodological Issues in Measuring the Uninsured." Proceedings of the Seventh Health Survey Research Methods Conference, Williamsburg, VA. 167–173. Available at: http://www.cdc.gov/nchs/data/hsrmc/hsrmc_7th_proceedings_1999.pdf (accessed on February 5, 2016).

- Pascale, J. 2001b. "The Role of Questionnaire Design in Medicaid Estimates: Results from an Experiment." Talk presented to the Washington Statistical Society, March 21, 2001.
- Pascale, J. 2004. "Medicaid and Medicare Reporting in Surveys: An Experiment on Order Effects and Program Definitions." Proceedings of the American Association for Public Opinion Research, American Statistical Association, May 13–16, 2004. pp. 4976–4983. Available at: <http://www.amstat.org/sections/SRMS/Proceedings/> (accessed on April 2, 2015).
- Pascale, J. 2008. "Measurement Error in Health Insurance Reporting." *Inquiry* 45(4): 422–437. doi: 10.5034/inquiryjrnl_45.04.422 Available at: <http://inq.sagepub.com/content/45/4/422.full.pdf+html> (accessed on February 4, 2015).
- Pascale, J. 2009a. "Findings from a Pretest of a New Approach to Measuring Health Insurance in the Current Population Survey." Paper prepared for the Federal Committee on Statistical Methodology Research Conference, November 2–4, 2009. Available at: https://fcsml.sites.usa.gov/files/2014/05/2009FCSM_Pascale_VIII-A.pdf (accessed on February 5, 2016).
- Pascale, J. 2009b. "Survey Measurement of Health Insurance Coverage: Cognitive Testing Results of Experimental Questions on Integrated Current and Calendar Year Coverage." Available through "Q-BANK," a database of questions and reports on cognitive testing maintained by the National Center for Health Statistics. Available at: http://www.nchs.gov/qbank/report/Pascale_Census_2009_HealthInsurance.pdf (accessed on February 5, 2016).
- Pascale, J. 2009c. "Health Insurance Measurement: A Synthesis of Cognitive Testing Results." Paper presented at the Questionnaire Evaluation Standards (QUEST) meeting, May 18–20, 2009, Bergen, Norway.
- Pascale, J. 2015. "Modernizing a Major Federal Government Survey: A Review of the Redesign of the Current Population Survey Health Insurance Questions. Study Series, Survey Methodology #2015–03. Available at: <https://www.census.gov/srd/papers/pdf/SSM2015-03.pdf> (accessed April 2016).
- Pascale, J., M. Boudreaux, and R. King. 2015. "Understanding the New Current Population Survey Health Insurance Questions." *Health Services Research* 51: 240–261. Doi: <http://dx.doi.org/10.1111/1475-6773.12312>.
- Pascale, J., J. Rodean, J. Leeman, C. Cosenza, and A. Schoua-Glusberg. 2013. "Preparing to Measure Health Coverage in Federal Surveys Post-Reform: Lessons from Massachusetts." *Inquiry: The Journal of Health Care Organization, Provision, and Financing* 50: 106–123. Doi: <http://dx.doi.org/10.1177/0046958013513679>.
- Pascale, J. 2011. "Findings from a Split-Ballot Experiment on a New Approach to Measuring Health Insurance in the Current Population Survey." Report prepared for the Center for Survey Measurement, US Census Bureau.
- Pascale, J., M.I. Roemer, and D.M. Resnick. 2009. "Medicaid Underreporting in the CPS: Results from a Record Check Study." *Public Opinion Quarterly* 73: 497–520. Available at <http://inq.sagepub.com/content/45/4/422.full.pdf+html> (accessed on April 2, 2015).
- Redline, C. 2013. "Clarifying Categorical Concepts in a Web Survey." In "Topics in Survey Measurement and Public Opinion." Special issue, *Public Opinion Quarterly* 77: 89–105.

- Research Project to Understand the Medicaid Undercount. 2008. "Phase II Research Results: Examining Discrepancies between the National Medicaid Statistical Information System (MSIS) and the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC)." Available at: https://www.census.gov/did/www/snacc/docs/SNACC_Phase_II_Full_Report.pdf (accessed on October 5, 2015).
- Resnick, D. 2013. "Microsimulation Support for Tax, Transfer & Health Insurance Policy Analysis Summary." Submitted to David S. Johnson & Charles Nelson, U.S. Department of Commerce, U.S. Census Bureau, and Don Oellerich, U.S. Department of Health and Human Services, ASPE, on October 3, 2013.
- Resnick, D., S. Love, C. Taeuber, and J. Staveley. 2004. "Analysis of ACS Food Stamp Program Participation Underestimate." Paper presented at the 2004 Joint Statistical Meeting, August 8–12, 2004, Toronto.
- Ringel, J.S. and J.A. Klerman. 2005. "Today or Last Year? How Do Interviewees Answer the CPS Health Insurance Questions?" RAND Labor and Population working paper series WR-288. Santa Monica, CA: RAND.
- Roman, A.M., A. Hauser, and A. Lischko. 2002. "Measurement of the Insured Population: The Massachusetts Experience." Paper presented at the 2002 Annual Meetings of the American Association for Public Opinion Research, May 16–19, 2002, St. Pete's Beach, FL.
- Rosenbach, M. and K. Lewis. 1998. "Estimates of Health Insurance Coverage in the Community Tracking Study and the Current Population Survey." Document no. PR98-54. Washington, DC: Mathematica Policy Research, Inc.
- Schaeffer, N.C. and S. Presser. 2003. "The Science of Asking Questions." *Annual Review of Sociology* 29: 65–88. Doi: <http://dx.doi.org/10.1146/annurev.soc.29.110702.110112>.
- Sudman, S., A. Fin, and L. Lannon. 1984. "The Use of Bounded Recall Procedures in Single Interviews." *Public Opinion Quarterly* 48: 520–524.
- Swartz, K. 1986. "Interpreting the Estimates from Four National Surveys of the Number of People Without Health Insurance." *Journal of Economic and Social Measurement* 14: 233–242.
- State Health Access Data Assistance Center (SHADAC). 2013. *Comparing Federal Government Surveys that Count the Uninsured*. State Health Access Data Center.
- U.S. Census Bureau. 2012. "America's Families and Living Arrangements: 2012: Average Number of People: AVG1. Average Number of People per Household, by Race and Hispanic Origin/1, Marital Status, Age, and Education of Householder: 2010." Available at: <https://www.census.gov/hhes/families/data/cps2012AVG.html> (accessed on October 5, 2015).
- U.S. Census Bureau. 2014. "Health Insurance in the United States: 2014 (P60-253), Detailed Tables, Table HI01. Health Insurance Coverage Status and Type of Coverage by Selected Characteristics: 2014." Available at: <http://www.census.gov/hhes/www/hlthins/data/index.html> (accessed on October 5, 2015).
- U.S. Census Bureau. 2015a. "Measuring Health Insurance Coverage with the Current Population Survey: A History of Improvement." Available at: http://www.census.gov/content/dam/Census/library/infographics/measuring_health_insurance.pdf (accessed on October 2, 2015).

- U.S. Census Bureau. 2015b. "About Health Insurance." Available at: <https://www.census.gov/hhes/www/hlthins/about/> (accessed on October 5, 2015).
- Willson, S. 2005. *Cognitive Interviewing Evaluation of the National Immunization Survey Insurance Module: Results of Fieldwork and Laboratory Interviews*. Unpublished report. Hyattsville, MD. National Center for Health Statistics.

Received September 2014

Revised February 2016

Accepted February 2016