

Measuring Health Insurance in the U.S.: A Case Study in Methodology

Joanne Pascale

National Centre for Social Research, UK

1. Introduction

Unlike most industrialized countries in the world, the United States does not have a system of universal health coverage. Most individuals are covered by either public or private insurance (or both), but a substantial number of people are left without coverage of any kind. Measuring and tracking the number of uninsured, therefore, has been a high priority for the public health research and policy community. Furthermore, among those who are insured, determining the source of their coverage has important implications for researchers and policymakers. Status and source of coverage, in conjunction with other survey data, are routinely used to address a number of research questions, such as: whether various public programs are reaching their target populations; why some of those eligible for public programs are not enrolled; differences between the insured and uninsured in health status and use of health services; and whether the quality of care varies by type of insurance. Finally, survey estimates are used to allocate federal funds for certain public programs (Davern, 2003).

Surveys play a key role in measuring the number of uninsured simply due to the lack of a centralized insurance system. That is, the only way to estimate the uninsured is to measure the insured through surveys, and then derive the number of uninsured as those not covered. However, there are several major surveys -- both governmental and non-governmental -- that measure health insurance and the estimates across these surveys vary substantially. According to the most recent Current Population Survey (CPS), which is the U.S. government's official source of estimates on the uninsured, there were 15.6% uninsured throughout the year in 2003 (DeNavas-Walt, Proctor and Mills, 2004), 69% with private coverage, and 27% with some type of government coverage (note that many individuals have coverage through more than one source). However, over the past decade or so, estimates of the uninsured across various surveys have ranged from a low of about 8% up to a high of almost 18% (Bennefield, 1996; Lewis, Ellwood & Czajka, 1998; Rosenbach & Lewis, 1998). And a recent comparison of estimates from the CPS and the Survey of Income and Program Participation (SIPP), a panel study designed to measure economic well-being, reveals more than a two-fold difference in 1998 -- 16.4% uninsured in the CPS, versus 8% uninsured in the SIPP (Bhandari, 2004).

The design features of these surveys vary considerably and many questions remain about the relationship between these features and the accuracy of the estimates. The objective of this research, broadly, has been to take a single research question (how many uninsured?) and systematically address the possible sources of measurement error using a range of methods and a series of iterative studies over time to learn more about the various survey design features and their relationship to the estimates. The ultimate goal is to gain some insights into which questionnaire may best measure the uninsured. The hope, furthermore, is that this general approach to assessing measurement error may be useful if applied to other topic areas.

The remainder of this paper is organised as follows. Section II presents a general description of the major surveys that measure health insurance, similarities and differences in their design

features and estimates, and some of the research that has attempted to explain these differences. Section III covers prior research on data quality, identification of the gaps in the research and some of the more fruitful areas for future research. In Section IV some of the highlights of original research conducted will be discussed, and finally Section V offers a discussion and conclusions.

2. Different Surveys, Different Design Features

In addition to the CPS and the SIPP discussed above, there are several other surveys that are important to the research community. Among them are the governmental Medical Expenditure Panel Survey (MEPS) and the National Health Interview Survey (NHIS), and the privately-sponsored Community Tracking Survey (CTS) and the National Survey of America's Families (NSAF). Each of these surveys uses its own set of design features, including sample design and size, mode, weighting and imputation and instrumentation. While any one of those features could contribute to variations in the estimates derived, the current research addresses only instrumentation. The main reason for this is simply to stay focused on one aspect of the survey design. However, an earlier comparative study examined four different surveys (CPS, NHIS and two earlier incarnations of the MEPS) and assessed the effects of four design features on the estimates: sample framework, weighting, imputation and instrumentation (Swartz, 1986). Sample framework and weighting were found to be virtually identical, and imputation proved too idiosyncratic and, in some cases, proprietary to assess. The author points to instrumentation as a main source of measurement error that likely explains some variation in the estimates.

2.1 General Approach: Source-by-Source versus Status/Source

In very broad terms, health insurance instruments aim to determine whether individuals are covered by certain types of insurance – private and public plans. Within private coverage, the categories of interest are job-based and directly purchased coverage; within public coverage the categories are Medicare (generally for the elderly), Medicaid (designed mainly for low-income), the Children's Health Insurance Program (CHIP, covering low-income children), and military plans. Note, however, that these are only broad categorizations; Medicaid and Medicare often cover overlapping populations. For example, though Medicaid is primarily designed to cover low-income people, it may also cover nursing home care for the elderly. And though Medicare's primary population group is the elderly, it may also cover some non-elderly disabled people. Furthermore, it is not uncommon for individuals to be covered by both Medicare and Medicaid.

Though all the surveys mentioned above share the same goal – to classify individuals as insured or uninsured, and among the insured to identify source of coverage – the instrumentation used for these purposes varies. In fact the range of subtle variations in wording is quite vast; therefore only the more fundamental differences in instrumentation are addressed here. First is the general approach of the question series. To capture coverage under each of these categories, the CPS uses a 'source-by-source' approach and simply asks a series of yes/no questions -- one on each of the eight plan types. The series is followed by a question to verify that those who said 'no' to the entire series is actually uninsured. Finally the uninsured are calculated as those who verify they have no coverage. A fundamentally different approach, employed by the NHIS and other surveys, is to first ask a question on overall status – covered or not. Among those who say 'yes' a follow-up question is asked to determine which type of plan the respondent has.

2.2 Household- versus Person-level Design

Another important design feature is whether questions are asked at the 'household' or the 'person'

level. In the household-level design, a 2-part set of questions is asked about each plan type. First a household-level screener is asked to determine if anyone in the household is covered by a particular plan type. If yes to the household-level screener, a follow-up question is asked to determine which household members are covered by that plan type. Then the next plan type is asked about in the same manner, and so on until all plan types are covered. The person-level approach, on the other hand, asks a complete set of questions on each plan type for the first person in the household, then the second person, and so on, until all household members are covered. So the series would read: ‘Is Person 1 covered by a job-based plan?’ ‘Is Person 1 covered by a directly-purchased plan?’ and so on, until the last plan type is asked about Person 1. Then the series would repeat itself for Person 2, and all other household members, until the last household member is asked about.

The methodological issues related to this design feature have to do with respondent fatigue and specificity of household members. The household-level design may be more ‘user-friendly’ and less burdensome, since it asks fewer questions overall (especially as the number of household members increases), but may result in underreporting because the respondent is not prompted with each individual household member by name. There is a risk that without this explicit prompt, the respondent may not adequately consider all household members when thinking about a particular plan type. The person-level, on the other hand, may generate more reports of coverage because it does ask about each household member by name, but in households with two or more people it is more tedious and burdensome. There is a risk that this burden may result in respondent fatigue. That is, after the entire series is repeated for the fourth or fifth household member, the respondent may tire of the series and simply say ‘no’ to all plan types just to speed the interview along. This may, in turn, cause underreporting. Very little is known from the empirical literature, however, regarding this design feature and its effects on reporting.

2.3. Reference Period

A third major design feature that has received a fair amount of attention in the literature is reference period. Some surveys, including the CPS, use a calendar year reference period, asking: ‘At any time during 2004, were you covered by [plan type]?’ With this design, the uninsured are defined as anyone not covered throughout the entire year. That is, even if someone were covered for as little as a day, and uninsured throughout the rest of the year, they would be defined as insured. The CTS and other surveys use a ‘current’ or ‘point-in-time’ reference period and simply ask: ‘Are you currently covered by [plan type].’ The uninsured are then defined as those without coverage at a particular point-in-time (i.e.: the day of the interview). Still other surveys use a reference period somewhere in-between. The SIPP, for example, uses a 4-month reference period, asking ‘At any time during the past 4 months were you covered by [plan type].’ If yes, month-by-month data on coverage is collected. The uninsured can then be defined in any number of ways: uninsured throughout an entire month, or a 4-month period, or data can be aggregated to get a calendar year estimate of those uninsured throughout the entire year. Measurement issues here have to do mainly with recall. Generally, for mundane topic areas like health insurance, the longer the reference period the more memory decay.

This type of memory decay may account for some discrepancies uncovered in a number of comparative studies that have examined estimates from current and calendar-year surveys. One set of these studies has aggregated year-long data from surveys with 3 or 4 month reference periods and compared them to estimates from a calendar-year survey. For example, the National Medical Care Expenditure Survey (NMCES, which is a precursor to the MEPS study), used a 3-month reference period, and data were aggregated to a calendar year and compared to the CPS.

Research showed the difference in the uninsured estimate to be striking: the CPS estimate of the uninsured was one-and-a-half times higher than the NMCES estimates (Swartz, 1986). Another study compared an estimate from an aggregated calendar year of the SIPP to the CPS calendar year estimate with similar results; the CPS uninsured estimate was almost two times higher than the aggregated SIPP estimate (Bennefield, 1996). And as mentioned above, a more recent study virtually replicated this finding; for the year 1998 the CPS estimate was 16.4%, versus 8% uninsured in the SIPP (Bhandari, 2004).

Another set of studies compared the estimates from surveys with calendar year and point-in-time reference periods. In this case, however, we would expect to see differences in the estimates because the definition of the uninsured is different. Specifically, the point-in-time design asks about coverage on the *day* of the interview, while the calendar year design asks about coverage on *any* day over the past year. In theory we should find more people who have some kind of coverage if we're asking about *any* day throughout the year than we if we ask about one *particular* day of the year. The calendar year design, then, should result in more reports of coverage than the point-in-time design, and hence we should see a lower rate of uninsured in the calendar year design. However, one study compared the CPS with the CTS, which employed a very similar design with the exception of reference period. After controlling for known differences between the surveys, the study found the uninsured estimates to be virtually identical across the two surveys (Rosenbach & Lewis, 1998). Though this was a very provocative study, it wasn't a completely 'clean' comparison since it could not possibly control for all the many subtle design differences across the two surveys.

3. Prior Research, Gaps and Agenda for Current Research

Given these variations in design features and estimates, and prior research, the objective of the current research was to assess what was known about data quality and, based on that, to carve out a research agenda to further explore gaps in the research. Again due to a lack of comprehensive records or any type of 'gold standard' of truth regarding coverage, this assessment turned to three general criteria. First was the extent to which the various designs had been evaluated and pretested. More specifically, had there been any type of pretesting involving the respondent that would indicate how well the respondent understood what was being asked? Did respondents have the knowledge to answer in the level of detail requested, particularly when asked to report for other household members? Second, had the survey questions been tested for validity? Finally, had they been assessed for reliability?

Regarding the question of respondent comprehension and knowledge, for the CPS approach there appeared to be no comprehensive cognitive testing on the overall series. This is probably due in large part to the fact that the series was developed and implemented before such testing was commonly available. In more recent years, however, there was some cognitive testing of two questions within the series – specifically the questions on the Children's Health Insurance Program (CHIP) and Medicaid. CHIP is a relatively new federal program, introduced in 1997, designed to cover children from low-income families who earn too much to qualify for Medicaid. Researchers found that respondents had a very difficult time recognizing the CHIP program name within the question series, and also had trouble distinguishing it from Medicaid. An unanticipated but perhaps not surprising finding was that some respondents confused Medicare with Medicaid, sometimes erroneously responding 'yes' to the first plan type in the series because it sounded similar to the actual plan they were covered by (Loomis, 2000). Researchers speculated that context effects could affect estimates in large production surveys.

There was also some cognitive research at the state level, in Massachusetts, on reporting of government plans. That research showed that respondents found it somewhat difficult to report whether they were covered, and that specifying the type of coverage they had was even more challenging (Roman, 2002). The authors concluded that respondents with government insurance generally do know they are covered, but if the questions on types of government coverage don't map on to their own perceptions of those plans, they say "no" to government insurance and end up saying "yes" to something else -- often private coverage -- depending on the order of the questions. As in the CHIP/Medicaid research, these authors speculated that context effects could be important.

Regarding qualitative research on the alternative approach to measuring health insurance used by NHIS – asking a 2-part series, first about status and then source of coverage – there had been two rounds of cognitive testing, in 1995 and again in 2002. The results were similar in both tests: the status question (covered or not) did not appear to pose major reporting problems but respondents had trouble categorizing their coverage in the 'source' question (Beatty and Schechter, 1998; Beatty et al, 2002).

Turning to validity and reliability, a literature review in 2000 took on this question for one survey in particular, the Behavioral Risk Factor Surveillance Survey, which is structured like the NHIS. The authors found no studies on reliability, and only one on validity. Results of the validity study showed that the status question (covered or not covered) resulted in reasonably valid data, but identifying particular source of coverage had poor validity (Nelson, 2001).

On the question of data quality, then, the literature seems to be pointing in the same direction: respondents can report status without too much difficulty but source is much more prone to error. The research agenda that emerged based on existing evidence, then, included both qualitative and quantitative work. Qualitative research involved departing from both the source-by-source approach and the status-source approach and redesigning the series, taking into account respondents' difficulty in reporting source. Another qualitative study was to conduct cognitive testing 'from scratch' on the CPS approach since it has never been done, even though the general design is in common use. On the quantitative side, split ballot experiments were designed to provide empirical data on suspected problem areas, such as reference period (calendar year vs. point-in-time), household- vs. person-level design, and context effects (e.g.: sequencing of Medicare and Medicaid items).

4. Current Research Findings

Following are highlights of findings produced under this general research agenda. First, with regard to qualitative research on the basic design, given that prior research suggests that respondents can answer status but have difficulty with the source, an alternative and perhaps more respondent-friendly approach may be to use a "funneling" technique. This approach would first ask a global, inclusive question about health insurance status, since that seemed unproblematic. But then, rather than going straight to a detailed question about the particular type of coverage, it would ask a very general question on the basic source of coverage, listing four categories: an employer, direct-purchase, the government, or something else. Then followup questions tailored to each of these four general sources would be asked to tease out the necessary detail, such as which type of government plan. Research along these lines was conducted in two phases of cognitive testing – first in 2001 and again in 2003.

The first round was promising in terms of the general approach – asking respondents to categorize their plan into one of the four basic types of coverage. But the second stage – asking tailored questions designed to gather more detail on each of the general sources – needed refinement (Pascale, 2001). More work was conducted to better specify the questions on each of these four paths and the funneling approach was tested again. The second round was much more promising, turning up no major problems but a series of small, manageable problems. For example, respondents who were employees of the federal government sometimes had difficulty choosing between the ‘employer’ and ‘government’ from the list of four basic source categories. Probes were built into the question, instructing interviewers to code this as employer-based coverage. Furthermore, for respondents who chose the government as their source, a followup question was asked to determine if this coverage was related to *employment* through the government in any way (versus public assistance) (Pascale, 2003). This redesigned series was then adapted for computer-assisted telephone interviewing (CATI) and experimentally tested alongside the CPS approach in a large-scale field study, called the 2004 Questionnaire Design Evaluation Research Survey, with roughly 1,000 cases per treatment. Results of that study are still pending. However, during trainings and debriefings with interviewers, no serious problems emerged. Furthermore, respondents were asked for their consent to link their survey data to administrative records, in order to facilitate a subsequent validation study. Finally, a subset of interviews was tape recorded for later behavior coding. These research projects are also still pending. Regarding qualitative work on the CPS, basic was cognitively testing was conducted in fall, 2004, and results from this testing is also still pending.

On the quantitative side, a series of split-ballot experiments was conducted over the past few years, beginning in 1999. The first experiment used the basic CPS design but experimentally manipulated the household- and person-level design. Results showed that overall the person-level design generated more reports of coverage than the household-level design, and this was driven entirely by more reports of job-based plans in the person-level treatment. Given that job-based plans are by far the most prevalent plan type, higher reporting of this plan type would greatly contribute to differences in the overall uninsured rate. Differences in reporting for the other plan types were generally small and mixed; for some plan types the person-level generated more coverage, and for other plan types the household-level design got more reports, and no differences were significantly different for the overall sample. However, when the sample was divided into ‘small’ households (1-3 people) and ‘large’ households (4 people or more) there was an interesting finding for Medicaid reporting. Among small households the person-level design resulted in an estimate of 6.2% of household members covered by Medicaid, while the household-level design estimate was almost twice that -- 11.5%, a statistically significant difference. This could be evidence of respondent fatigue discussed earlier. That is, in the person-level design, once respondents have listened to the entire series of eight questions for the first few people in the household, they may tire of the series and fail to attend to or report for later people in the household.

Results of the reference period experiment were consistent with prior research. In comparing calendar year with current reference period, the calendar year design should, in theory, pick up more reports of coverage since it’s offering a much longer time period in which to be covered. Therefore the uninsured rate under the calendar year design should be lower than under the current design. Results from the split-ballot experiment, however, showed the uninsured rate to be exactly the same across both designs – 12.0%. This replicates the CPS/CTS comparison discussed above, but offers more ‘clean’ evidence since the experiment controls for all other survey design

features and conditions (sampling method, interviewer training, mode, response rates, etc.). Findings suggest that indeed there is some kind of measurement error associated with reference period, since there should necessarily be a gap in the calendar year and current estimates of uninsured. As with all studies like this that lack some type of validation source, however, it is not entirely clear which reference period design is the ‘culprit’. Given evidence from the survey methods literature on memory decay over time, and the increased complexity of the calendar year design (‘at any time during 2004 were you...’ versus ‘are you...’), it is likely that the calendar year design is problematic.

Finally, on the issue of context effects, in the CPS source-by-source design, private plans are asked first, followed by public plans, in the following order: job-based, directly-purchased, coverage from someone outside the household, Medicare, Medicaid, SCHIP, military, and other plans. In a split-ballot experiment, this order was maintained for one treatment, and for the other treatment the overall order was preserved but the Medicare/Medicaid order was flipped. The three types of private plans were still asked about first, but then the order was as follows: Medicaid, SCHIP, Medicare, military and other plans. Results also showed no main effects for the overall sample – specifically no evidence of Medicaid underreporting, which came as comforting news to those concerned with this type of measurement error. However, there was some evidence of double-reporting; some respondents who were covered by Medicaid seemed to report that coverage at both the Medicare *and* the Medicaid question. More specifically, among poor households, there were more reports of dual coverage (being on both Medicaid and Medicare) when Medicare was asked first than when Medicaid was asked first (Pascale, 2004). There are many ways to interpret this finding. One possibility is that observations from cognitive testing are manifesting in the production data. That is, respondents covered by Medicaid who get their coverage confused may mistakenly say ‘yes’ to the Medicare item, when it comes first, and then also report their Medicaid coverage when that item is asked, resulting in overreporting of Medicare and correct reporting of Medicaid.

5. Conclusions

Like all fields of study, surveys and survey research are evolving. Techniques for designing good questionnaires are being refined over time, as are the methods used for evaluating the quality of data they produce. In the case of health insurance measurement, standard instruments were generally developed before pretesting methods such as cognitive testing were a routine part of survey data production. The result is that there are now generally two paths of research. One is post-hoc testing and evaluation of the most common instruments used to estimate the uninsured. The other is to develop an instrument that takes full advantage of what is known from this post-hoc analysis but also incorporates more of the pretesting and evaluation methodologies that more explicitly involve the respondent’s perspective in the interview process.

The current research represents attempts in this direction. Perhaps one of the most robust findings, suggested by secondary data analysis and validated by a split-ballot experiment, is that there is some type of measurement error associated with the reference period. Qualitative research and a validity study suggest that respondents can report on their status but have difficulty reporting source under both the source-by-source design and the status-source design. Preliminary qualitative evidence on a redesigned ‘funneling’ approach suggests that respondents may have an easier time reporting their coverage – at least in terms of broad categories. Pending studies on this redesign will shed more light on its efficacy. Findings on the person- versus household-level design were rather mixed. Until further research replicates the findings, it may be premature to

conclude that the person-level design is superior. Qualitative and quantitative research on ordering effects suggests that indeed there may be misreporting based on whether Medicaid or Medicare is asked first, but the error may not be underreporting but overreporting.

Clearly there are still many unanswered questions on both the particulars of health insurance measurement and the more general reporting issues surrounding it. The hope is that this research could serve as a kind of 'landscape' of the issues to date by (1) pulling together what is known about data quality of health insurance survey data (2) contributing some original findings to the debates (3) generating new questions, and (4) suggesting fruitful avenues for future research. On this last point, this may involve studies that attempt to bring in a validation piece, such as record check studies.

References

- Beatty, Paul, Barbara Wilson, Kristen Miller, Alfredo Calvillo, Karen Whitaker, (2002), NHIS Insurance Module and additional questions on Work, Income, and Health Care Utilization, Unpublished report, April, 2002
- Beatty, Paul, and Schechter, S. (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, No. 26, p. 12-14
- Bennefield, R. (1996). A comparative analysis of health insurance coverage estimates: Data from CPS and SIPP. Presentation at the 1996 Joint Statistical Meetings of the American Statistical Association.
- Bhandari, Shailesh, People with Health Insurance: A Comparison of Estimates from Two Surveys, *Survey of Income and Program Participation Report No. 243*, June 8, 2004.
- Davern, Michael, Lynn A. Blewett, Boris Bershadsky, Kathleen Thiede Call, and Todd Rockwood. 2003. State Variation in SCHIP Allocations: How Much Is There, What Are Its Sources, and Can It Be Reduced. *Inquiry*. 40(2): 184-197.
- DeNavas-Walt, Carmen, Bernadette D. Proctor, Robert J. Mills, "Income, Poverty and Health Insurance Coverage in the United States: 2003," Washington, DC, August 2004.
- Lewis, K., Ellwood, M., & Czajka, J. (1998). Counting the uninsured: A review of the literature. Washington, D.C., Mathematica Policy Research.
- Loomis, Laura (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, Unpublished Census Bureau report, Washington, D.C.
- Nelson, David, D. Holtzman, J. Bolen, C.A. Stanwyck, K.A. Mack (2001), Reliability and validity of measures from the Behavioral Risk Factor Surveillance System (BRFSS), p. S3-S42, Supplementum (Special Issue) no 1, 2001 *Journal Social and Preventive Medicine: International Journal of Public Health*
- Pascale, J. (2001), "SIPP 2001 Methods Panel Wave 1/Replicate 2 Recommendations Based on Cognitive Testing Results: Health." Report submitted to SIPP Methods Panel Team, Jan. 9, 2001.

Pascale, J. (2003), Questionnaire Design Experimental Research Survey (QDERS) 2004: Cognitive Testing Results on Health Insurance Questions. *Unpublished Census Bureau Report*, November 5, 2003, Washington, D.C.

Pascale, J. (2004), Medicaid and Medicare Reporting in Surveys: An Experiment on Order Effects and Program Definitions, *Paper presented to the American Association for Public Opinion Research Annual Meeting*, May, 2004, Phoenix, AZ

Roman, Anthony M., A. Hauser, 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*. St. Pete's Beach, Fla., May, 2002.

Rosenbach, M. & Lewis, K. (1998). Estimates of Health Insurance Coverage in the Community Tracking Study and the Current Population Survey. Washington, D.C., Mathematica Policy Research, Document No. PR98-54.

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.