Special Section

Federal Health Data: National and State Level Uses and Issues

Toward a More Reliable Federal Survey for Tracking Health Insurance Coverage and Access

Genevieve Kenney, John Holahan, and Len Nichols

Objective. Examination of the extent to which federal surveys provide the data needed to estimate the coverage/cost impacts of policy alternatives to address the problem of uninsurance.

Data Sources/Study Setting. Assessment of the major federal household surveys that regularly provide information on health insurance and access to care based on an examination of each survey instrument and related survey documentation and the methodological literature.

Study Design. Identification of the data needed to address key policy questions on insurance coverage, assessment of how well existing surveys meet this need, definition of the critical elements of an ideal survey, and examination of the potential for building on existing surveys.

Data Collection/Extraction Methods. Collection and critical assessment of pertinent survey documentation and methodological studies.

Principal Findings. While all the federal surveys examined provide valuable information, the information available to guide key policy decisions still has major gaps. Issues include measurement of insurance coverage and critical content gaps, inadequate sample sizes to support precise state and substate estimates, considerable delays between data collection and availability, and concerns about response rates and item nonresponse. Our assessment is that the Current Population Survey (CPS) and the National Health Interview Survey could be most readily modified to address these issues.

Conclusions. The vast resources devoted to health care and the magnitude of the uninsurance problem make it critical that we have a reliable source for tracking health care and coverage at the national and state levels and for major local areas. It is plausible that this could be more cost effectively done by building on existing surveys than by designing and fielding a new one, but further research is needed to make a definitive judgment. At a minimum, the health insurance information collected on the CPS should be revised to address existing measurement problems.

Key Words. Health insurance coverage, uninsured rates, estimates and measurement of health insurance coverage

According to the most recent estimates from the Current Population Survey (CPS), approximately 45.5 million nonelderly Americans lack health insurance—long a major national issue that promises to remain so (Denavas-Walt, Proctor, and Lee 2005). The implications of lack of insurance for access to care and the population's health have been documented in numerous studies (Institute of Medicine 2002; Hadley 2003), and alternatives to address the problem have been proposed from across the political spectrum. But, unfortunately, the absence of reliable and consistent data on insurance coverage and uninsurance makes it very difficult to give policymakers valid and defensible estimates of the cost and coverage implications of policy proposals, at both the national and state level. Indeed, frustration over the flawed nature of the state-level data that are available from existing surveys has led a large number of states to develop their own surveys (Blewett and Davern 2006). These parallel efforts are problematic because they are costly and because they produce competing estimates that have their own methodological limitations.

And this is in spite of the fact that five publicly funded national household surveys currently provide estimates of the uninsured: the CPS, the National Health Interview Survey (NHIS), the Medical Expenditure Panel Survey (MEPS), the Survey of Income and Program Participation (SIPP), and the Behavioral Risk Factor Surveillance Survey (BRFSS).

In debates about the merits of improving our national data, cost is frequently cited as a major barrier. But in the context of our current health system, which costs \$1.7 trillion a year and includes over \$40 billion in uncompensated care for the uninsured (Hadley and Holahan 2003), even the cost of a new survey (\$200 million or so) would be a minor investment. Moreover, modifications to existing federal surveys could plausibly achieve the essential improvements at lower additional cost.

Each of the five surveys we have listed is designed to address different public policy questions, all of which are important. But each also has limits in

Address correspondence to Genevieve Kenney, Ph.D., The Urban Institute, 2100 M Street, NW, Washington, DC 20037. John Holahan, Ph.D., is with The Urban Institute, NW, Washington, DC. Len Nichols, Ph.D., is with the New America Foundation, NW, Washington, DC.

Table 1: Uninsured Estimates from National Surveys

Survey	Most Recent Year	Method of Estimate		
		Uninsured For Full Year	At the Time of the Survey	Uninsured at Any Time during the Year
Current Population Survey (CPS), Census	2003	45.0 million (15.6%)	NA	NA
Medical Expenditure Panel Survey (MEPS), HHS*	2002	32.4 million (11.5%)	47.3 million (16.6%)** (2003)	62.6 million (22.3%)
National Health Interview Survey (NHIS), HHS	2003	28.8 million (10.0%)	42.0 million (14.6%)	51.4 million (17.9%)
Survey of Income and Program Participation (SIPP), Census	2001	18.9 million (6.8%)	38.7 million** (14.0%)	66.5 million (24.0%)

Notes:

*Nonelderly.

**Estimate for March 2001.

NA, Survey does not capture this dimension; and point in time refers to the period from January 1 through the interview date, which is 5 months on average.

Source: O'Grady, M.J. "The Uninsured: Policy and Data Issues." Presentation at the American Enterprise Institute, Washington, DC, April 28, 2005.

its ability to collect the information we need for analyzing the problem of uninsurance and the implications of different approaches to addressing it. Considering them together, five weaknesses stand out.

No Agreement on Coverage Estimates. As can be seen in Table 1, there is no consensus among the federal surveys on how many Americans are uninsured—at a point in time, for a year, or for some time during a year.¹ This is problematic because the health implications for the long-term uninsured are likely quite different than for those who spend a brief time between insurance coverages—as are the reasons why they are without insurance and what it would take (i.e., cost) to cover them effectively. Take, for example, the estimated 45.5 million people without insurance mentioned above. This comes from the CPS, by far the most widely quoted database on the extent of the uninsurance problem, and is reported as the number of Americans who lacked health insurance “last year,” interpreted as the *whole of the past year*. This estimate differs by more than a factor of two from the comparable estimate derived from the SIPP (Table 1, Peterson and Devere 2002). It also is very similar to the 42.0 million estimate from the NHIS of the number uninsured *at a point in time* (which all observers agree is substantially higher than the number who are uninsured for a whole year). Even the two longitudinal surveys (the MEPS and SIPP) produce very different estimates of the uninsured *at a point in time* (47.3 versus 38.7 million) and of the uninsured *for a full year* (32.4 million versus 18.9 million).

Lack of State- and Local-Level Estimates. None of the five surveys can support a full set of precise state-level estimates of coverage or related access and use measures. As states are responsible for much of the decision making that affects the uninsured, this gap has serious consequences for health policy making. It is not necessary that precise state-level estimates be available for all states annually; for most states, having such precise state-level estimates available on a regular 2–4 year cycle would be a real step forward. This would allow states to support their ongoing policy debates without incurring the large expenditures required to finance their own surveys.

Insufficient Information on Type and Use of Coverage. Estimates of health insurance coverage by source of coverage (especially for the two most important sources, employer sponsored and Medicaid) also vary considerably across surveys. Moreover, most of these surveys do not cover important dimensions of the benefit packages to which people have access and use, such as total and out-of-pocket premium costs, other cost sharing, and covered services. In addition, not all of the major household surveys have information on access to and use of health care services. These are crucial for determining

policy-relevant factors such as adequacy and affordability of coverage and for assessing access for the uninsured.

Insufficient Information for Eligibility Determination. The surveys do not always collect income information that is linkable to specific family members or include detail on income sources—crucial for simulating eligibility for public health insurance programs. Nor do they always collect detailed household roster information necessary for identifying health insurance units.

Extensive Time Lags Until Data Release. Data are most useful for direct policy analysis when they are available in a timely manner. There are inherent tradeoffs that need to be recognized between both the complexities and comprehensiveness of surveys and the lag in data availability. But with the pace of change in our health system, policy analysts need prompt release of national and selected state-level estimates and data for quick analysis of proposed reform packages, in addition to the longer lag times needed for release of public use files of sufficient quality to support in-depth longer-range study using longitudinal data on, e.g., how coverage is changing for children who are eligible for Medicaid.

THE SURVEY REQUIREMENTS THAT MUST BE MET

Four survey requirements must be met if these data weaknesses are to be remedied: the validity of the estimates must be universally agreed to be without bias; the sample size must be large enough to produce precise national and state estimates; there must be appropriate content with requisite detail and flexibility; and data release must be timely. We discuss each briefly in turn.

Validity of the Estimates

The two major issues with respect to validity are nonresponse and under-coverage. To address the former problem, high response rates are needed. The two major types of survey frame—the population from which the sample is drawn—are (1) area frames, for which the sample is drawn from a current list of residential addresses, with in-person or a combination of in-person (e.g., for persons without phones) and telephone interviewing, versus (2) random digit dialing (RDD) frames with telephone interviewing. Area frames have much higher response rates than RDD frames. In addition, response rates on RDD surveys have been declining, raising concerns about the validity of estimates of change over time (Massey, O'Connor, and Krotki 1997). Even with substantial resources invested in mailings, prepayment, and incentive payments, it does

not appear that response rates in RDD surveys can approach those achieved with in-person surveys (Brick et al. 1999).

The fundamental concern with low response rates is that nonresponders are concentrated in subgroups that differ systematically from responders in ways that cannot be adjusted for by the sample weights. This problem is aggravated in the case of telephone surveys, because the 3–4 percent of Americans estimated to live in households without a telephone (2000 Census) are known to be different from those with telephones, for example, in being disproportionately poor, residing in the South, and being uninsured (Davern et al. 2001). Weighting strategies have been developed to adjust estimates based on RDD surveys that do not include nontelephone households (Blumberg et al. 2004). But the complexity of such strategies can lead to delays in data release. More fundamentally, weighting strategies can introduce biases of their own in estimates of the uninsured, particularly for the subgroups disproportionately likely to be without a phone. This problem can be alleviated to some extent by including a supplemental area frame, which provides at least some information on households without phones. However, to the extent that the sample weights do not fully adjust for coverage and nonresponse issues, estimates of the number and composition of the uninsured will be biased.

Response rates are also affected by whether the survey is cross-sectional or longitudinal. Longitudinal surveys, following the same households over multiple survey periods, are critical for estimating key behavioral parameters related to the dynamics of insurance coverage and expenditure and use patterns and for estimating duration of coverage. But sample attrition and relatively low response rates associated with longitudinal surveys make the development of weights more complex and time consuming than for most cross-sectional surveys, resulting in data release delays and raising concerns about bias. Moreover, by conducting multiple interviews with the same household over the course of a year, longitudinal surveys involve higher costs for a given sample size. This makes a cross-sectional survey more likely to be a cost-effective way of achieving the goals of a survey designed to track cross-sectional changes over time.

Precision of the Estimates and Scale

It is critical that the survey produce precise estimates at a point in time for the nation, and periodically for states and at least some localities and markets. It is also desirable that such a survey permits detection of changes in insurance coverage and related measures over time. What size change needs to be

detected? At the national level, ideally, the survey should be capable of detecting a one-percentage-point change in insurance coverage. This may sound small, but a one-percentage-point decline in the number of adults with employer-sponsored coverage reflects an absolute decline of close to two million adults. To detect a change of this magnitude from one year to the next would require a national effective sample size of up to 40,000 adults, and the same number of children for a similar change in the coverage of children.

Achieving comparable precision at the state level (i.e., detecting small changes in uninsured rates at the state level) would require a national sample of millions of observations. In addition, to produce valid state-level estimates it is essential that the survey include samples that are representative of the residents in each state.

Content Detail and Flexibility

The ideal survey designed to delineate the major dimensions of the uninsurance problem would cover the following areas: a wide range of detail about health insurance coverage or lack of it; detailed information on income; demographic and socioeconomic characteristics; health and functional status; household composition and relationships; and health care access, use, and out-of-pocket spending. In addition to adequate detail on these topics, it is critical that the content be flexible enough for adaptation to changes in the policy and market environment. For example, when a new program such as the State Children's Health Insurance Program (SCHIP) is implemented, insurance questions need to be modified promptly to capture the new coverage source. The two highest priority areas—insurance coverage and income—are discussed immediately below.

Insurance Coverage. While there are multiple ways to define the size of the uninsured population—i.e., the number uninsured at a point-in-time versus the number uninsured for a full year, each serving a useful purpose from a policy perspective—there is a lack of consensus on how best to measure a given concept. A number of methodologies are currently in use to measure insurance coverage (Lewis, Ellwood, and Czajka 1998; Short 2001). Although there is no definitive methodological research as yet to ascertain which one produces the truest measure of insurance coverage for a given concept, some of the approaches have limitations that are generally recognized:

- Asking only a single question, as is done on the BRFSS is problematic. It is critical to know how the *patterns* of coverage are changing

over time. Moreover, relying on a single question will likely lead to biased estimates of the extent to which people have any type of insurance coverage.

- Asking whether a person has any of a number of different types of coverage and identifying the number of uninsured as the residual tends to overstate the number of uninsured (Rajan, Zuckerman, and Brennan 2000; Nelson and Mills 2001). For this reason, several surveys now ask a question confirming uninsurance—which can affect the distribution of coverage across different types of insurance and the characteristics of the uninsured, as well as a revision of the total number of uninsured. Interestingly, a confirmation question has led to differing degrees of change in the rate of uninsurance (Rajan, Zuckerman, and Brennan 2000), plausibly depending on the sequencing of the coverage questions.
- There is general agreement that measurement error is likely to be lower when respondents are asked about current coverage than when asked to recall coverage over a longer period (Short 2001).
- Asking only about current coverage is insufficient. As already noted, the duration of uninsurance is likely to be related to the reasons for uninsurance. In addition, as health care access and use are generally measured over a period of time, there is a potential disconnect between those measures and current coverage. It is likely that the duration of a specific type of coverage and patterns of coverage over time are best measured in a longitudinal survey. As can be seen in Table 1, the two federal longitudinal surveys (MEPS and SIPP) produce estimates of the number uninsured at any point during a year that are over 10 million higher than the estimate produced by the NHIS, which is cross-sectional. While cross-sectional surveys that include questions on duration of coverage and prior coverage patterns can provide useful information, more methodological research is needed to determine how they can provide more accurate estimates of these important measures.
- It is critical that measures of insurance coverage include reliable measures of enrollment in public health insurance programs (such as Medicaid and SCHIP). It is known that most surveys yield lower estimates than the programs' administrative data (Bennefield 1996; Lewis, Ellwood, and Czajka 1998; Blumberg and Cynamon 2001; Call et al. 2001).² Administrative data may have their own biases, but

existing research certainly suggests that introducing state-specific names and including multiple questions increases the public coverage that is reported. But more methodological research is needed to improve the quality of the information provided on household surveys about enrollment in public programs.

- Addressing nonresponse to specific questions (item nonresponse) is likely to produce sounder estimates of insurance coverage than ignoring the problem and weighting up the observations with the coverage information that is available. However, the approach for imputing missing coverage data may introduce new problems (Davern et al. 2004); some model-based approaches to imputation offer the potential of preserving unbiased estimates (Little and Rubin 1987), which could be particularly important for state or local level estimates.
- Asking about the coverage of multiple persons in the household versus asking about each specific individual leads to different results (Blumberg et al. 2004). Additional research is required to ascertain which way the reporting bias runs.

Income. Having sound income information is key to assessing eligibility for public programs and to understanding how insurance coverage is changing for children and adults in different income groups. Here are the major issues to be addressed:

- It is well known that relying on a single question to determine household or family income leads to measurement error (see, for example, Moeller and Mathiowetz 1994; Davern 2003). Although a single question can lead some respondents to overstate and others to understate (Cantor and Wang 2000), this approach appears to lead, on balance, to an understatement of household income.
- Given the importance of income information in clarifying health insurance policy issues, an adequate survey for this purpose must have a battery of questions that both includes the amount of income derived from a number of different possible sources (Moore, Stinson, and Welniak 2000) and associates that income amount with a particular individual in the household.
- The broader the range of potential sources, the larger is the total income amount captured.
- Associating a given income source with a particular person is essential for identifying the income of health insurance or eligibility units

within the household which is crucial for simulating the cost and other impacts of different reform proposals (Dubay and Kenney 2000; Short 2001; Davern et al. 2003).

- It is important that income measures track with the key time periods over which insurance coverage and health care access and use are being measured which requires asking about past as well as current income.
- It is important to understand in the context of collecting income information that a significant amount of nonresponse is inevitable in any area regarded as a private and sensitive matter. Addressing missing income data with imputation is key to developing sound income measurement, no matter how detailed the income questions may be.

Timely Data Release

To be most useful and relevant from a policy perspective, estimates need to be released within a short time-frame, for example months, following the data collection period, which itself needs to be brief. It is also important that public use files with geographic identifiers be available as soon thereafter as possible, as the release of public use files greatly increases the extent to which the survey data will be used and analyzed. About 6 months seems to be the shortest feasible period between data collection and release for surveys as large and complex as the ones we are discussing here.

HOW THE FIVE FEDERAL SURVEYS STACK UP ON THE ABOVE CRITERIA

Each of the five federal household surveys we have listed meets important data needs. But each is also deficient with respect to one or more of the criteria listed in the previous section as crucial for definitive data on the uninsurance problem. (More detail on these and other surveys is on the Economic Research Initiative on the Uninsured website and is available at http://www.umich.edu/~eriu/researcher/resources_datamain.html.)

The CPS (Annual Social and Economic Supplement)

The primary purpose of the CPS is to collect monthly employment and unemployment statistics (U.S. Census Bureau 2003). A secondary purpose is to collect information on the demographic status of the population, including

health insurance coverage. Information on health insurance coverage is included as part of the Annual Social and Economic Supplement, which is fielded in March of each year. The CPS is a very large cross-sectional survey, based on the entire civilian noninstitutionalized population, with a sample size of 99,986 households (216,000 individuals). This large-scale survey is based on an area frame (which reduces the bias suffered by telephone surveys that cannot, by definition, reach households without land lines) that provides representative samples in every state. It yields the most precise national estimates available on an annual basis. It is also large enough to yield precise state-specific estimates on an annual basis for larger states and reasonably precise estimates with 2 years of data for a subset of additional states (Blewett et al. 2004). Its response rate is high—85 percent in the March 2003 survey—adding credibility to its estimates. The CPS is conducted with both telephone and in-person interviews.

The CPS contains detailed information on household composition and on demographic and socioeconomic characteristics. It also collects comprehensive income information, including earned and unearned income, from multiple sources for each person in the household. It collects income information at the person level, which has the great advantage of allowing users to group income into health insurance units. The income information has one weakness, however. It is collected only for the previous year and not for the current period, which makes it vulnerable to recall problems, as well as being less informative for assessing current eligibility for public coverage.

A more major weakness for our purposes is the way the CPS asks for health insurance information, which leads to confusion in interpretation of the estimates. Its basic question on this topic is: “At any time in (the previous year) was anyone in the household covered by (a particular type of insurance)?” The survey names each of several possible types of private and public insurance. For each positive response on the various types of insurance it includes, the survey then asks which household member(s) was covered by that particular kind of insurance. The CPS thus obtains insurance status for each person individually, which is a strength. It also recently added a question verifying that each individual for whom no coverage is reported is in fact uninsured (Short 2001), which has reduced the estimated number of the uninsured by 9 percent (Nelson and Mills 2001).

Whether the information on the CPS yielded by this form of questioning in fact captures insurance coverage for the past year is the subject of great debate because the answer has major effects on policy estimates of the costs of reform, among other things (Lewis, Ellwood, and Czajka 1998; Czajka and

Lewis 1999). Here is the essence of the debate: The way the CPS makes estimates of the number of uninsured is by counting anyone who did not respond positively to having some form of coverage (taking into account the responses provided at the verification question) as uninsured for the whole of the past year. Doubt about the accuracy of the CPS estimates of those lacking insurance all year arises because those estimates are more consistent with other surveys' estimates of uninsurance *at a point in time* than with the other estimates of uninsurance *over a whole year*. This inconsistency has led researchers to debate whether CPS respondents are either (a) responding to the CPS by actually providing their current coverage at the time of the survey or providing an indication of their typical or average coverage situation during the past year, or instead (b) responding correctly about the past year, but with some recall error because of the long reference period, in which case the similarity between the CPS estimates and the point in time estimates from other surveys is mere coincidence (Swartz 1986; Short 2001; U.S. CBO 2003; Denavas-Walt, Proctor, and Lee 2005). Indeed, new data from Maryland (Eberly 2005) indicate that many respondents did not report enrollment in public coverage in the past year if they were no longer enrolled at the time of the survey.

An additional issue with the health insurance estimates produced by the CPS is that item nonresponse is over 10 percent for the health insurance questions (Davern et al. 2001), much higher than item nonresponse for insurance questions on the NHIS, which is below 1 percent. Moreover, questions have been raised about the approach used for imputing health insurance data on the CPS, particularly for producing state-level estimates (Davern et al. 2004).

The content of the CPS is lacking in a number of areas. First, the method of collecting uninsurance information yields no direct information on coverage status at the time of the survey or of how long people may have been without health insurance. Second, the CPS only periodically asks whether employers offer coverage to workers or dependents and, if so, whether employers pay all, part, or none of the premium as part of the February supplement. This is useful information, but it is fielded only sporadically (in 2001, for example, and not again until February 2005). In addition, it is only the subsample interviewed in both February and March that has information on both insurance coverage and employer offers. Finally, the CPS has only a single question on health status (i.e., whether an individual is in excellent, very good, good, fair, or poor health) and no information on health care utilization, access, or spending.

The CPS is the model among household surveys in terms of the lag between data collection and release of estimates. Information on health insurance coverage is collected during March every year, and the data are released about 6 months later (in 2004, the lag was just 5 months)—including estimates and public use files, based on the full sample, with state identifiers and imputed data on income and insurance coverage. The short field period (which is just a month) likely contributes to the speedy release of data and estimates. This timely release is a major reason for its wide use (Short 2001) and is plausibly key to why the CPS is commonly viewed as the survey of record for national and state-level estimates of insurance coverage, despite the measurement issues detailed above.

NHIS

The NHIS is the main source of information on the health and health care of the U.S. population (U.S. Department of Health and Human Services, Centers for Disease Control and Prevention 2005). This cross-sectional survey, fielded annually over the course of the year, is used to monitor trends in illness and disability, and health care access and utilization. Like the CPS, it is a nationally representative sample of the U.S. civilian noninstitutionalized population, based on an area-sampling frame. In 2002, the NHIS covered 38,000 households and 93,386 individuals. The NHIS includes each state but is not designed to be state representative; thus only national and regional data are made routinely available. State estimates are possible for certain large states where the distribution of primary sampling units is state representative, though no state identifiers are available on public use files (Blewett et al. 2004). Future data releases will reportedly include estimates for about 10 (presumably large) states. The survey is conducted in person (also avoiding the likely biases associated with telephone surveys) and has a high household response rate (89.6 percent in 2002.)

The NHIS has detailed information on health status, the presence of functional limitations, and the presence of chronic conditions. It also has detailed information on utilization and access, which covers a wide range of measures—including usual source of care, unmet need, hospital stays, and doctor visits. Most of this information is asked for the previous 12-month period (with additional information collected on a subset of measures using a 2-week reference period), leading to a temporal mismatch between some of the utilization and access information and the measurement of insurance coverage, which focuses on coverage at the time of the survey.

The NHIS provides substantial demographic and socioeconomic information. However, its income information is somewhat limited. Respondents are asked if they or other family members received income from each of a list of sources in the prior calendar year (which includes earnings for each adult family member reported as “working”) and total family income. However, the amounts attributed to specific sources are not exhaustive—i.e., they do not add up to the “total.” Thus, the available income information is total family income with flags indicating income from each of several sources. Because income data are not collected separately for each family member (except earnings for members reported as working), it is not possible to create income for health insurance units without imputation. In addition, income data cannot be aligned with the current insurance coverage information, as no information is collected on current income.

The health insurance questions ask whether anyone is covered by a health insurance plan at the time of the interview (Short 2001); if there is a positive response, more detailed questions are asked about what type of insurance that individual has. As the NHIS is fielded continuously throughout the year, it provides an average weekly measure or rolling average of insurance coverage for the year (Czajka and Lewis 1999). Additional questions are asked to ascertain whether the employer offered insurance coverage and whether individuals lacked coverage at any time in the past 12 months as well as the number of months the person was uninsured (U.S CBO 2003). Thus, the NHIS allows for estimates of the number of uninsured at a point in time, all year, or ever uninsured during the past 12 months, although the latter two estimates may be subject to error because of recall bias, which may be inherent in duration-dependent estimates that are derived from cross-sectional surveys.

The NHIS also contains a verification question to check whether the respondents for which no insurance information was given are in fact uninsured (although the addition of this question led to a much smaller change in the uninsured rate than it did for the CPS). In the middle of 2004, the NHIS added a follow-up question for those who were not classified as having any of the different types of health insurance coverage listed in the survey to address possible underreporting of Medicaid coverage among the nonelderly population (Cohen and Martinez 2005). The impact of this follow-up question was not conclusive because it did not have a statistically significant effect on the coverage estimates; however, it could increase Medicaid coverage rates by as much as one percentage point for children and by a much smaller amount for nonelderly adults, with concomitant decreases in uninsurance.

The NHIS does not allow a portrayal of the *types* of coverage that insured individuals might have had at different times over the past 12 months. Furthermore, it collects only very limited information on out-of-pocket spending and none on health care spending by insurers that pays for the care used by individual.

In recent years, through its Early Release Program, the NHIS has been releasing preliminary health insurance coverage estimates by income (but with no imputations for missing data) 6–12 months after the data were collected. In December 2004, for example, the NHIS released preliminary coverage estimates based on data collected between January and June 2004 and in September 2005, the NHIS released partial year estimates, based on data collected between January and March 2005. However, the most recent public use files with imputed income information, released in March 2005, came only after a substantial time lag (based on data collected over the course of 2003), which is almost 2 years after the data were collected, on average.

The MEPS-HC

The MEPS is designed to provide comprehensive information about health care use, spending, insurance coverage, and sources of payment (Agency for Healthcare Research and Quality 2003). It is a longitudinal panel survey of the entire civilian noninstitutionalized population, with a sample drawn from NHIS respondents in the previous year. This allows person-linked data to be obtained from both surveys. The MEPS is conducted in person (avoiding the bias from telephone surveys) and the response rate in 2000 was 65.8 percent (the combined response rate for the NHIS and the MEPS). Because of the modest sample size (37,000 individuals in 2002 and 15,000 households), only national and regional level data are available (Blewett et al. 2004), though estimates have been released for the 10 largest states (Machlin and Sommers 2005) and research is being conducted aimed at producing estimates for a broader set of states (Sommers 2005). The survey is conducted annually and is designed to produce valid cross-sectional estimates each year. MEPS panels are interviewed every 3–6 months with a total of five interviews over a 2-year period.

Given its primary purpose, the MEPS provides detailed information on utilization, access, and expenditures over the same periods for which income and insurance coverage information is collected (Agency for Healthcare Research and Quality 1998a, b; Short 2001). It also contains a wide range of information on health status, functional limitations, and chronic conditions.

Its expenditure estimates include spending not only by the respondent but also by others, (e.g., insurers that pay for the care used by the individual and his family).

The MEPS collects all the requisite household roster and demographic and socioeconomic information. It also collects detailed income data by income source for each member of the household or family, thus allowing for creation of income for health insurance units. The MEPS collects earnings and related data each round of the survey.

The MEPS also collects detailed insurance coverage information. It asks whether anyone in the family has been covered by a particular type of insurance at any time during the reference period, i.e., the past 3–6 months. If there is a positive response, the survey then asks who was covered by that particular type of insurance. Interpreting estimates that are specific to the reference period can be confusing, however, because respondents are reporting for periods of different lengths. For those who report coverage during the reference period, additional questions are asked about which months an individual was covered, about coverage at the time of the interview, and about coverage during the months covered by the reference period. Through this series of questions the MEPS can produce estimates of the uninsured at a point in time, whether individuals are ever uninsured during the past year, and whether they were uninsured all year (Short 2001). Because of the way the MEPS is fielded, income estimates can also be derived to correspond to the different insurance estimates. The MEPS provides information on offers of coverage to workers and how much, if anything an employer contributes to the employee premiums. In addition, because MEPS also fields an employer survey, there is a potential to enhance the information collected on the household survey with more detailed information on the nature of the insurance coverage offered by the employer.

MEPS releases multiple public use files for different purposes. The most relevant for the purposes of tracking insurance coverage is the release of “point-in-time” files, which are released a little over a year after the data were collected. In July 2004, for example, MEPS released estimates and a subset of the full data for a data collection period that took place a year and several months earlier (February–May 2003).

SIPP

The SIPP is a longitudinal survey that collects data on income, labor force characteristics, public program participation, and demographics. The survey

collects data on the entire civilian noninstitutionalized population and is based on an area frame. Some in-person interviews are conducted, with the remaining interviews conducted by telephone. The sample size is 37,000 households (the 2001 panel). Data are collected for the same panel over a period of 2–4 years, with panel individuals interviewed every 4 months. Given its modest sample size, only national and regional estimates can be made (Blewett et al. 2004). In addition, its particular panel structure, unlike the MEPS, is not designed to produce sound annual estimates as the sample is not replenished with a fresh cross-section each year of the panel. While the response rate for the first few waves of the panel is high (for example, the response rate for the SIPP in FY 2001 was 86.7 percent), response rates decline in subsequent years of the SIPP panel because of the cumulative effects of sample attrition.

The SIPP collects detailed information on demographic and socioeconomic conditions as well as income information for the previous 4 months. Income information by source is collected in considerable detail for each family member, thus allowing for the creation of income for health insurance units.

Insurance coverage information is asked about for each family member. For those who report coverage, additional questions are asked about coverage at the time of the interview, coverage during the month of the interview (U.S. CBO 2003), and coverage during each of the previous 3 months. The survey allows for uninsurance estimates at a point in time, all year, or any time during the year. Information is collected on employer offers to workers and to dependents and whether an employer contributes all, part, or none of the premium. The SIPP does not have a verification question, but does have questions that restate a person's insurance status, which may allow for some correction for the underreporting that occurs when there is no type of summary check.

The survey collects limited information on utilization and access to care for the previous 12 months. Out-of-pocket spending data are asked during some waves of the survey but not all, as is information on health status and on the presence of functional limitations. The data collected from 2001–2003 as part of the 2001 SIPP panel were released during 2004.

Behavioral Risk Factor Surveillance System (BRFSS)

The BRFSS is designed to monitor trends for the adult population in preventable health conditions and risk behaviors that are related to injury and chronic and infectious diseases (U.S. Department of Health and Human Services, Centers for Disease Control and Prevention 2004). It is an annual

survey that collects data on 150,000 adults. State-specific estimates are available and estimates for substate areas are also available for some states. The survey is an RDD survey, conducted by telephone only, with the resulting bias through noncoverage of nontelephone households being addressed through weighting procedures. Concerns remain, however, about whether this weighting still leaves some undersampling of low-income households (Blewett et al. 2004). States choose their own survey vendors, which could result in variable quality among states—a concern that gains added plausibility given the wide range of state-specific response rates. In 2002, for example, rates varied from 25 to 79 percent depending on state (with the overall median response rate being 45 percent). While reliance on an RDD sample frame raises concerns about the quality of the BRFSS data, it offers flexibility in terms of content that has been exploited to inform emerging policy issues (e.g., in Fall 2004, new questions were added in light of the flu vaccine shortage). The BRFSS also collects annual income information, but only for the adult in the household who is interviewed and with no detail on income source.

Insurance information collected by the BFRSS is also quite limited, with only one broad question about coverage at the time of the interview, “Do you have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, or government plans such as Medicare?” This yields a measure of uninsurance at the time of the interview, but there is no information on the length of uninsurance, employer offers, employer contributions to premiums, or out-of-pocket costs. The information on utilization and access is also limited and cannot—because it is asked for the past 12 months—be linked directly to the information collected on insurance coverage. The BRFSS includes questions on health status and some information on chronic conditions. Data are collected over the course of a year and generally, the data files are released 6 months after the completion of data collection.

What Is the Potential for Building on Existing Surveys?

As indicated above, each of these five federal household surveys meets important data needs, but none, as currently configured, provides reliable and timely state and national estimates on coverage, access, and cost that are needed to inform policy. The two existing federal health surveys that hold the most promise for providing more reliable tracking information on coverage at the state and national level, in our judgment, are the CPS and the NHIS—both are based on area frames, have high-response rates, include nontelephone households, and are fielded annually as cross-sectional surveys. However, as

indicated above, both have shortcomings with respect to their health insurance content, in addition, the NHIS has inadequate detail on income, while the CPS lacks information on health care access and use.

Building off the MEPS is appealing on its face, because it is the one existing federal survey that covers all the content needed. And clearly the MEPS should be retained as it provides critical information on health care expenditures and duration of insurance coverage, for example that is not available from any other existing federal survey. However, its sample size is too small to support precise state-level estimates and to expand its size would require first expanding the NHIS.³ In addition, MEPS is an inefficient mechanism for obtaining timely cross-sectional estimates, as its current panel design would make it very costly to expand it sufficiently to produce state-level estimates, because of the high costs of maintaining a longitudinal panel.

The SIPP is not a reasonable candidate because it is not fielded annually and even the years covered by SIPP panels do not include a fresh cross-section each year. This, because of panel attrition, greatly limits its usefulness for providing cross-sectional estimates. Moreover, its measurement of insurance coverage, which currently lacks an explicit verification question, would have to be retooled, and additional content would be needed on health care access, use, and spending and on health status.

The BRFSS, with its low-response rates and exclusive focus on telephone households makes it inherently vulnerable to concerns about the reliability of the estimates. Moreover, the survey would have to be broadened to include children and the content would also have to be expanded, as it has very limited information on income and household relationships, insurance coverage, and health care access and spending.

The following describes how the CPS and the NHIS could be modified to meet the criteria we specified above.

Building on the CPS. Although both would have to be modified substantially, the CPS would require less change to its sample design than the NHIS, as the CPS already has a state-representative sample frame, which is essential for providing reliable state-level estimates. In fact, the CPS is the only federal health insurance survey that releases state-level estimates for both children and adults. However, as indicated above, the CPS currently falls short in terms of total sample size for producing precise estimates for many states, and it has limited ability to estimate coverage or anything else at the substate level.

What scale is necessary to achieve a reasonable set of objectives? As noted above, detection of a 1-percentage-point change in insurance coverage

from 1 year to the next with a high degree of precision would require a national effective sample size of 40,000 for each group (i.e., adults and children). The CPS is currently large enough for this purpose. But achieving comparable precision at the state level (i.e., 40,000 children and 40,000 adults in each state) would require millions of sample observations. Even achieving a total nominal sample of 5,000 children in each state annually would require a major expansion of the CPS—by a factor of 3.6 over the expanded sample implemented in 2001. It should be noted that a significant sample size expansion to the CPS would likely require an increase in the number of primary sampling units in many states as well.

One option would be to obtain high levels of precision annually for states that have relatively large populations and settle for less precision for smaller states. Large nationally representative surveys naturally include large samples for a number of high-population states—as happens currently with the CPS. The California, New York, and Texas samples of children (ages 0–18) number over 3,000, for example; those for Florida, Illinois, Michigan, Ohio, and Pennsylvania number over 2,000—compared with under 1,000 in the 21 smallest states.

One scenario would be for the 10 largest states to have a sample size of 10,000 each annually, and for each of the other states plus Washington, DC to have annual samples that are, say, one-half to one-third that size. This would allow fairly small changes to be detected annually for the population in the 10 largest states, which include about 54 percent of the nation's population. For the smaller states, annual estimates would be less precise, as would estimates of change over time. Combining 2 or 3 years of data, however, would increase precision. This approach would lead to a total effective sample size of 250,000–320,000. Achieving an effective sample size of the magnitude would involve a major expansion of the CPS given that the current nominal CPS sample size of adults under age 65 is 195,343 (which yields an effective sample size of about half that size, because of the complex sample design).

An alternative design option for the smaller states would be to include larger annual samples for subsets of states on a rotating basis, such that every 2–4 years, every state would have enough precision to assess coverage patterns for key subgroups within the state for that year. This would add complexity to the sample design and make it more difficult to assess cross-state differences on an annual basis. But it would provide smaller states with a better opportunity to gauge periodically the size and nature of the health care problems in their state.

With large state-specific samples and an appropriate mix of primary sampling units, it should also be possible to produce fairly precise estimates for some local areas, particularly for large or otherwise important MSAs, either annually or by combining several years of data together. Assessing local variation cross-sectionally and over time is important as very few market areas for health insurance and health care are statewide. In addition, supplementing state and local sample sizes with estimates derived from model-based small area estimation techniques could help address information gaps for small states and local areas (Citro, Cohen, and Kalton 1998).

For the CPS to become the undisputed source of valid health insurance estimates, even if its sample size is increased as suggested above, it is essential that its content be modified. At a minimum, the health insurance information would have to be changed to permit reliable measurement of insurance coverage at the time of the survey and in the preceding year (as outlined in the section on measurement of insurance coverage). Clearly, such a change in measurement would likely compromise the continuity of the health insurance estimates over time (although a phase-in of the new questions could reduce that problem). However, without such a change, debates will continue about the fundamental meaning of the estimates.

Information on access to employer-sponsored insurance coverage (the nature and cost of the coverage that is available) and on current income (which has been available only periodically on the February supplement to the CPS) would also have to be added. Finally, targeted information on access, use, out-of-pocket spending, functional limitations, and chronic conditions would need to be added as well.

Building on the NHIS. Building on the NHIS would require even greater changes to the sample than for the CPS. First, the NHIS is not currently designed to produce state-representative estimates. Second, the NHIS sample would have to be increased a great deal to achieve the desired precision. Its sample size, for example, is less than half as large as that of the CPS. Third, in order for the NHIS to become the survey of record in terms of insurance coverage estimates, the early NHIS release would need to be retooled to include final, fully imputed estimates of both insurance coverage and income and would need to be based on more reliable income measures. In order for the NHIS to be used externally for tracking insurance coverage to the same extent as the CPS, public use files would also need to be produced more

quickly and state identifiers would need to be more accessible (whereas now they are only available at the research data center). Perhaps a compromise would be to release a file that has geographic identifiers but omits sensitive health data in order to address confidentiality concerns.

In addition, the fielding of the NHIS over the course of a year (compared with the CPS which collects the insurance coverage data in a single month) makes it inevitable that there will be a greater lag between when the data are collected and when they are released, even if the release occurs at the same point after the fielding is completed. Achieving a speedier release of coverage estimates based on the full NHIS sample would likely require a shorter field period, which could be at odds with monitoring health issues over the course of a year. Overall, the NHIS content covers most of the key data items, but its content would need to be changed in several areas—with more complete/reliable information needed on both past year's and current income, out-of-pocket spending, and past year's insurance coverage.

CONCLUSION

With all the resources devoted to health care and the magnitude of the uninsured problem in this country, it is critical that we have a reliable source of data for tracking health care access and coverage at the national, state, and (a sample of) local levels. But despite the crucial need for such data and the importance of the issue, identifying additional resources necessary to develop such a data set is bound to be difficult and controversial. Given that each of the existing surveys now contributes real value added, it is important that their essential features be preserved. Moreover, it would not be possible to create a single survey that addresses all the functions of the current surveys.

We have assumed in our discussion that building on existing surveys would be less expensive than designing a complete survey from scratch. This is a plausible assumption. But we obviously have not done the necessary calculations to be sure that it is the most cost-effective solution to the problem of building a definitive state and national database for assessing health insurance coverage and access to care.

We recognize that making this kind of decision requires answers to a number of outstanding questions. In particular, we would need to know the cost and technical implications of expanding either the CPS or the NHIS in the ways we have proposed. However, even if the CPS and the NHIS remain

at their current sizes, there would be real gains associated with addressing the content issues in both surveys. On the CPS, this would include improving the measurement of insurance coverage (i.e., including a measure of current insurance coverage) and adding measures of health status, access, and use. On the NHIS, this would include improving the information collected on income, out-of-pocket spending on health care, and on past year's insurance coverage.

In addition, we need to know how the costs of expanding and modifying the CPS or the NHIS compare with the cost and technical implications of developing and fielding a new survey that meets the criteria we propose, what cost-reducing modifications in existing surveys a new survey would allow, and what savings would be generated through these modifications. Here are examples of some of the questions that would need to be researched and answered: Could the sample size in the March CPS Supplement be reduced if a new health insurance survey were to be initiated or if the NHIS (or another federal survey) sample size were expanded? How much would be lost if the MEPS panel were shortened to just a year? If the CPS health insurance questions were improved, would there still be a need for NHIS to release general health insurance estimates? In addition, more information is needed on the potential for making greater use of statistical techniques to fill some of the gaps with existing data sources. For example, what could be gained if there were greater use of statistical matching across the existing surveys or if there was greater reliance on small area estimation techniques to measure key outcomes for small states and local areas?

In this review, we have focused on the federal surveys that currently measure insurance coverage. However, the American Community Survey (ACS), which planned to sample three million households nationwide in 2005, could be modified to include questions on health insurance coverage and related topics (currently, it collects information that draws almost exclusively from the Census Long Form). Given the scale of this ongoing effort and the potential for developing annual estimates for areas of over 65,000 inhabitants (and the ability to develop estimates for smaller areas based on 3 and 4 years of data) at low marginal cost, it makes sense to explore the feasibility of at least expanding the content of the ACS to incorporate key information on health insurance coverage at a minimum. At the same time, however, it will be important that any new estimates derived from the ACS complement existing estimates and not create more confusion about the extent and nature of the uninsured problem in this country. Moreover, it is very unlikely that the ACS could become the key health care tracking survey, given that so much

additional information would need to be added on health care access, use, and spending.

We believe that modifications to the CPS or the NHIS, or alternatively a new survey, could provide us with great benefits—though at significant costs. These costs could be reduced if existing surveys were appropriately modified after carefully consideration. In particular, current federal data collection efforts involve redundancy, as evidenced by the competing health insurance estimates put out by the CPS and the NHIS, which introduce confusion into debate on the uninsurance issue. Removing this redundancy could reduce the marginal costs of the new effort we believe is necessary to provide us with the insurance and access information the nation needs to address the major health care coverage problems it faces. Moreover, whether or not modifications are made to existing surveys or a new survey effort is launched, more methodological research is needed, particularly regarding how best to measure current and past year's insurance coverage, in the context of both cross-sectional and longitudinal surveys. Without such research and without a clearer delineation of the roles of the various federal surveys in providing these various coverage estimates, we will continue to have confusion over the extent and nature of the uninsurance problem facing this country.

ACKNOWLEDGMENTS

The authors appreciate: the suggestions and advice of Jim Baumgartner, Jessica Banthin, Lynn Blewett, John Czajka, Michael Davern, Amy Davidoff, Jonathan Gruber, Ed Hunter, Adam Safir, Tom Selden, Stephen Zuckerman, and two anonymous referees; the provision of Table 1 by Michael O'Grady; the editorial assistance of Felicity Skidmore; and the research assistance of Marie Wang, Jamie Rubenstein, and Justin Yee.

All the opinions expressed in the manuscript are those of the authors and do not reflect those of The Urban Institute or its funders.

NOTES

1. No estimate is included in the table from the BRFSS as it does not provide information on the insurance coverage of children.
2. Research currently underway in a number of states should provide more insights about why administrative totals do not align with survey estimates (Call 2005).

3. More research is needed to understand the implications of the measurement differences between the MEPS and the NHIS with respect to the point-in-time estimates. It is understandable why the past year measures differ between the two surveys as the panel nature of the MEPS likely increases the accuracy of the information collected on past year's coverage over what is obtained on the NHIS. However, the point-in-time estimates also differ between the two surveys and it is not clear a priori which estimate is more accurate. The MEPS was designed to provide complete information on coverage for the different reference periods, which leads to a more complex approach to determining current insurance coverage compared with the NHIS, which may have some bearing on why the two surveys lead to different estimates of current coverage.

REFERENCES

- Agency for Healthcare Research and Quality (AHRQ). 1998a. "MEPS Household Component (HC) Survey Instrument: Rounds 1-5; Health Insurance (HX) Section" [accessed on May 20, 2005]. Rockville, MD: AHRQ. Available at http://www.meps.ahrq.gov/Data_Pub/Questionnaires/questhc5/HX-p1r5.pdf
- . 1998b. "MEPS Household Component (HC) Survey Instrument: Rounds 1-5: Income (IN) Section" [accessed May 20, 2005]. Available at http://www.meps.ahrq.gov/Data_Pub/Questionnaires/questhc5/IN-p1r5.pdf
- . 2003. "MEPS HC-050: 2000 Full Year Consolidated Data File" [accessed May 20, 2005]. Available at <http://www.meps.ahrq.gov/puffiles/h50/h50doc.htm#C31SampDesign>
- Bennefield, R. L. 1996. *A Comparative Analysis of Health Insurance Coverage Estimates: Data from CPS and SIPP*. Paper presented at the Joint Statistical Meetings, Chicago, American Statistical Association.
- Blewett, L. A., and M. Davern. 2006 "State Decisions to Supplement Federal Data on State Health Insurance Coverage Estimates ." *Health Services Research* DOI: 10.1111/j.1475-6773.2006.00543.x
- Blewett, L. A., M. B. Good, K. T. Call, and M. Davern. 2004. "Monitoring the Uninsured: A State Policy Perspective." *Journal of Health Politics, Policy, and Law* 29 (1): 107–45.
- Blumberg, S. J., and M. L. Cynamon. 2001. "Misreporting Medicaid Enrollment: Results of Three Studies Linking Telephone Surveys to State Administrative Records." In *Seventh Conference on Health Survey Research Methods*, edited by M. Cynamon and R. Kulka, pp. 189–95. Hyattsville, MD: National Center for Health Statistics.
- 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." *National Center for Health Statistics. Vital Health Statistics* 2 (136): 1–38.

- Brick, J.M., G. Shapiro, I. Flores-Cervantes, D. Ferraro, and T. Strickler. 1999. "1997 NSAF Snapshot Survey Weights Report No. 3." *Assessing the New Federalism*, NSAF Methodology Reports. Washington, DC: Urban Institute.
- Call, K. T. 2005. *Survey and Administrative Data Sources of the Medicaid Undercount*, Presentation at State Health Access Data Assistance Center Spring Workshop, "Survey and Administrative Data Sources of the Medicaid Undercount," Washington, DC.
- Call, K. T., G. Davidson, A. S. Sommers, R. Feldman, P. Farseth, and T. Rockwood. 2001/2002. "Uncovering the Missing Medicaid Cases and Assessing Their Bias for Estimates of the Uninsured." *Inquiry* 38 (4): 396–408.
- Cantor, D., and K. Wang. 2000. "Correlates of Measurement Error When Screening on Poverty Status for a Random-Digit Dial Survey." 1997 Technical Papers: NSAF Methodology Report Number 16 Washington, DC: Urban Institute.
- Citro, C. F., M. L. Cohen, and G. Kalton. 1998. "Small-Area Estimates of School-Age Children in Poverty." Interim Report 2: Evaluation of Revised 1993 County Estimates for Title I Allocations. Washington, DC, National Academy Press.
- Czajka, J., and K. Lewis. 1999. *Using National Survey Data to Analyze Children's Health Insurance Coverage: An Assessment of Issues*. Washington, DC: Mathematica Policy Research, Inc.
- Davern, M. 2003. *Measuring Income in Health Insurance Surveys*. Minneapolis: University of Minnesota, SHADAC Survey Guidelines Series.
- Davern, M., L. A. Blewett, B. Berushadsky, and N. Arnold. 2004. "Missing the Mark? Possible Imputation Bias in the Current Population Survey's State Income and Health Insurance Coverage Estimates." *Journal of Official Statistics* 20: 519–49.
- Davern, M., L. Blewett, B. Berushadsky, and N. Arnold. 2001. *Possible Bias in the Census Bureau's State Income and Health Insurance Estimates*. Washington, DC: State Health Access Data Assistance Center. Presentation to HHES, US Census Bureau.
- Davern, M., G. Davidson, T. Beebe, K. T. Call, H. Rodin, T. Rockwood, and L. Blewett. 2003. *Are the Relationships between Health Insurance Coverage and Key Economic Demographic Characteristics Survey Dependent? A Comparison of the Behavioral Risk Factor Surveillance System and the Current Population Survey*. Minneapolis: University of Minnesota, State Health Access Data Assistance Center (SHADAC).
- Denavas-Walt, C., B. D. Proctor, and C. H. Lee. 2005. *Income, Poverty, and Health Insurance Coverage in the United States: 2004*, Current Population Reports P60-229. Washington, DC: US Census Bureau.
- Dubay, L., and G. Kenney. 2000. "Assessing SCHIP Effect Using Household Survey Data: Promises and Pitfalls." *Health Services Research* 35 (5, part III): 112–27.
- Eberly, T. 2005. "Maryland CPS Undercount Survey." Presentation at State Health Access Data Assistance Center Spring Workshop "Survey and Administrative Data Sources of the Medicaid Undercount," Washington, DC.

- Hadley, J. 2003. "Sicker and Poorer—The Consequences of Being Uninsured." *Medical Care Research and Review* 60 (2): 3S–75S.
- Hadley, J., and J. Holahan. 2003. "How Much Medical Care Do the Uninsured Use, and Who Pays for It?" *Health Affairs Web Exclusive* (February 12) W3-66-81
- Institute of Medicine. 2002. *Care without Coverage, Too Little, Too Late*. Washington, DC: National Academies Press.
- Lewis, K., M. Ellwood, and J. L. Czajka. 1998. "Counting the Uninsured: A Review of the Literature. *Assessing the New Federalism*, Occasional Paper No. 8.." Washington, DC: Urban Institute.
- Little, R. J. A., and D. B. Rubin. 1987. *Statistical Analysis with Missing Data*. New York: John Wiley.
- Machlin, S. R., and J. P. Sommers. 2005. *Estimates of Health Care Expenditures for the 10 Largest States, 2002. Statistical Brief # 69..* Rockville, MD: Agency for Healthcare Research and Quality.
- Massey, J. T., D. O'Connor, and K. Krotki. 1997. "Response Rates in Random Digit Dialing (RDD) Telephone Surveys." In *Proceedings of the 1997 American Statistical Association Section on Survey Research Methods, Anaheim, August 1997*, pp. 707–12. Alexandria, VA: American Statistical Association.
- Moeller, J. F., and N. A. Mathiowetz. 1994. "Problems of Screening for Poverty Status." *Journal of Official Statistics* 10 (3): 327–37.
- Moore, J., L. L. Stinson, and E. J. Welniak Jr. 2000. "Income Measurement Error in Surveys." *Journal of Official Statistics* 16 (4): 331–61.
- Nelson, C. T., and R. J. Mills. 2001. *The March CPS Health Insurance Verification Question and Its Effect on Estimates of the Uninsured*. Washington, DC: U.S. Census Bureau.
- Peterson, C. L., and C. Devere. 2002. "Health Insurance: Federal Data Sources for Analyses of the Uninsured." Congressional Research Service, Report to Congress.
- Rajan, S., S. Zuckerman, and N. J. Brennan. 2000. "Confirming Insurance Coverage in a Telephone Survey: Evidence from the National Survey of America's Families." *Inquiry* 37 (3): 317–27.
- Short, P. F. 2001. "Counting and Characterizing the Uninsured." Working Paper 2. Ann Arbor, MI, University of Michigan, Economic Research Initiative on the Uninsured (ERIU).
- Sommers, J. 2005. "Producing State Estimates with the Medical Expenditure Panel Survey—Household Component." Working Paper No. 05011. Rockville, MD, Agency for Healthcare Research and Quality.
- 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 (3): 233–42.
- U.S. Census Bureau. 2003. "Current Population Survey, 2003 Annual Social and Economic (ASEC) Supplement" [accessed May 20, 2005]. Available at <http://www.census.gov/apsd/techdoc/cps/cpsmar03.pdf>.
- U.S. Congressional Budget Office (CBO). 2003. *How Many People Lack Health Insurance and for How Long?* Washington, DC: CBO.

- U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. 2004. "2004 BRFSS Overview" [accessed March 28, 2006]. Available at http://www.cdc.gov/brfss/technical_infodata/surveydata/2004/overview_04.rtf
- . 2005. "2004 National Health Interview Survey (NHIS) Public Use Data Release: NHIS Survey Description" [accessed March 28, 2006]. Available at <http://www.cdc.gov/nchs/data/nhis/srvydesc.pdf>