© Health Research and Educational Trust
DOI: 10.1111/j.1475-6773.2008.00941.x
RESEARCH ARTICLE

An Examination of the Medicaid Undercount in the Current Population Survey: Preliminary Results from Record Linking

Michael Davern, Jacob Alex Klerman, David K. Baugh, Kathleen Thiede Call, and George D. Greenberg

Objective. To assess reasons why survey estimates of Medicaid enrollment are 43 percent lower than raw Medicaid program enrollment counts (i.e., "Medicaid undercount").
Data Sources. Linked 2000–2002 Medicaid Statistical Information System (MSIS) and the 2001–2002 Current Population Survey (CPS).
Data Collection Methods. Centers for Medicare and Medicaid Services provided the Census Bureau with its MSIS file. The Census Bureau linked the MSIS to the CPS data within its secure data analysis facilities.
Study Design. We analyzed how often Medicaid enrollees incorrectly answer the CPS health insurance item and imperfect concept alignment (e.g., inclusion in the MSIS of people who are not included in the CPS sample frame and people who were enrolled in Medicaid in more than one state during the year).
Principal Findings. The extent to which the Medicaid enrollee data were adjusted for imperfect concept alignment reduces the raw Medicaid undercount considerably (by 12 percentage points). However, survey response errors play an even larger role with 43 percent of Medicaid enrollees answering the CPS as though they were not enrolled and 17 percent reported being uninsured.
Conclusions. The CPS is widely used for health policy analysis but is a poor measure of Medicaid enrollment at any time during the year because many people who are enrolled in Medicaid fail to report it and may be incorrectly coded as being uninsured. This discrepancy should be considered when using the CPS for policy research.

Key Words. Medicaid undercount, MSIS, CPS-ASEC, survey measurement error, Medicaid

Survey estimates of public program enrollment are substantially lower than estimates of program enrollment compiled from administrative data for Medicaid, Temporary Assistance for Needy Families, and Food Stamps (C. Taeuber, D. Resnick, S. Love, J. Staveley, P. Wilde, and R. Larson,

unpublished data; Lynch et al. 2007; Call et al. 2008). This discordance is particularly apparent for Medicaid and has become known as the "Medicaid undercount" (Lewis, Elwood, and Czajka 1998; Klerman, Ringel, and Roth 2005; Call et al. 2008; Davern et al. 2008). The crude Medicaid undercount in the 2001 Current Population Survey (CPS)'s Annual Social and Economic Supplement (ASEC) was 42 percent, corresponding to calendar year 2000; in the 2002 CPS it was 43 percent, corresponding to calendar year 2001.[1]

This large Medicaid undercount in the CPS is particularly problematic because the CPS is widely used for official and unofficial health policy research purposes at the national and state level (Blewett et al. 2004). At the national level, CPS estimates are used by statute in the allocation of State Children's Health Insurance Program (SCHIP) funds to states (Davern et al. 2003). In addition, the Congressional Budget Office uses CPS-based estimates to "score" (i.e., estimate the cost of) legislation (Glied, Remler, and Zivin 2002). The CPS data are also used by state health policy analysts to examine the potential cost and impact of state-level health reform legislation and to report to the federal government on their progress toward insuring low-income uninsured children through SCHIP and other efforts (Blewett and Davern 2006). Unofficially the CPS is widely used by the academic and policy research community to evaluate health policy reforms and to estimate policy-relevant populations within each state, such as the number of people who are eligible for but not enrolled in public health insurance coverage (Blewett et al. 2004).

These uses of the CPS data emphasize the importance of an improved understanding of the Medicaid undercount in the CPS. Toward this end, this paper reports preliminary results from a project that linked MSIS Medicaid enrollment data to CPS survey data. The U.S. Census Bureau constructed files and performed tabulations that allow us to break the undercount into two components: (1) MSIS counts of people outside the CPS sampling frame and (2) survey response errors among those in the CPS sample. The resulting

---

Address correspondence to Michael Davern, Ph.D., Assistant Professor, Division of Health Policy and Management, School of Public Health, SHADAC—University of Minnesota, 2221 University Ave SE, Ste. 345, Minneapolis, MN 55414; e-mail: daver004@umn.edu. Jacob Alex Klerman, M.A., is with Abt Associates, Cambridge, MA. David K. Baugh, M.A., Senior Technical Advisor, is with Centers for Medicare & Medicaid Services, Office of Research, Development and Information, Baltimore, MD. Kathleen Thiede Call, Ph.D., Associate Professor, is with Division of Health Policy and Management, School of Public Health, SHADAC—University of Minnesota, Minneapolis, MN. George D. Greenberg, Ph.D., is with U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, Office of Health Policy, Washington, DC.

analysis presented in this paper provides insight into the contributions of these two components of the undercount, although an exact accounting is still not possible.

Data, Methods, and Analysis Plan

Our analysis uses the 2000–2002 Medicaid Statistical Information System (MSIS) data from the Centers for Medicare and Medicaid Services (CMS), the 2001 and 2002 CPS[2] survey data (reporting on health insurance coverage in calendar years 2000 and 2001), and linking procedures of the Census Bureau.

The Survey Data

The CPS is a monthly household survey of noninstitutional dwelling units. Its primary purpose is to generate the official monthly estimates of the unemployment rate. Individuals residing within selected dwelling units are interviewed according to a rolling panel design—the same four consecutive months in two successive calendar years. Interviews are primarily conducted in person in the first month that a dwelling unit is in the sample and via telephone or in person thereafter. Proxy responses are allowed; one member of the household generally responds for the entire household (U.S. Census Bureau 2002a).

While this monthly CPS survey only collects information sufficient to estimate the unemployment rate, the ASEC appends to the monthly CPS household interview detailed questions about income, employment, and program participation during the previous calendar year. Most of the ASEC interviews occur in March with some additional interviews occurring in February and April (U.S. Census Bureau 2002a; Davern et al. 2003). The 2001 and 2002 ASECs that we analyze here both had response rates of 84 percent (U.S. Census Bureau 2002b, 2003).

Since 1980, the ASEC has included a module on health insurance coverage. Consistent with the reference period for the income and labor force questions in the ASEC, the health insurance questions (including the Medicaid question) refer to the entire previous calendar year. This is in contrast to most other surveys that collect information on health insurance at the time of the interview (some also ask about coverage during the past year as well).

The basic ASEC health insurance question is structured to ask the household respondent if s/he or anyone else in the household had one of several different types of insurance coverage at any point during the last year.

After ascertaining that a specific type of coverage is operative for the household, a follow-up question is asked about who else in the household was enrolled. The specific types of health insurance coverage included in the household level screen portion of the interview are as follows: an employer or union plan; self-purchased insurance; someone outside the home provides coverage for anyone in the home; Medicare; Medicaid; SCHIP, other state-specific public health insurance programs; an item combining the VA, Military Health Care, and Indian Health Service; and a final "other" insurance please specify.[3] For any individual in the household with no affirmative answer for any of these types of health insurance, the household respondent is asked an uninsurance verification question: "I have recorded that (READ NAMES) were not covered by a health plan at any time in YEAR. Is that correct?" For both the "other" type of coverage and the verification question, a follow-up question for affirmative response allows the respondent to choose from the fifteen types of coverage (of which Medicaid is a possible answer).

In our analysis, we code all Medicaid responses as being "reported Medicaid" whether the Medicaid response came from the verification question, the "other public" health insurance question, or the Medicaid-specific survey item itself. However, the person needed to respond "Medicaid" or the state program name for Medicaid and not something else (e.g., SCHIP) to be included. We code all other types of coverage, including SCHIP, as other (non-Medicaid) types of health insurance coverage.

The Administrative Data

The MSIS data include an eligibility record for each individual enrolled in Medicaid. The Balanced Budget Act of 1997 mandated that states submit detailed individual-level Medicaid enrollment data to CMS. States are required to prepare data in a format specified by CMS; the data are then edited and cleaned by a CMS contractor and anomalies are noted in an appendix on the CMS website (CMS 2007a). Detailed information about the data element requirements is available from MSIS Tape Specification and Data Dictionary (CMS 2007b). For each recorded person, the MSIS data include the number of days enrolled by month, as well as various descriptors of coverage status and type. For our research, we use the 2000–2002 calendar year MSIS data files. Consistent with the CPS reference period, we count anyone who the MSIS indicated was enrolled with full-benefit Medicaid or enrolled in restricted Medicaid benefits for pregnancy services[4] for at least 1 day in the CPS calendar year reference period as having "full-benefit Medicaid" coverage

(calendar years 2000 and 2001). People enrolled during the year in SCHIP only and not Medicaid in the MSIS are not considered "full-benefit Medicaid" enrollees. We did this because SCHIP is not consistently submitted by the states for inclusion into the MSIS.

Linking the Administrative and Survey Data

For this project, CMS provided the Census Bureau with a version of its MSIS files for 2000–2002 that included an SSN. Within its secure data analysis facilities, the Census Bureau validated the SSN on the MSIS file and replaced the validated SSNs on the MSIS with the Person Identification Key (PIK), an internal Census identifier that represents a unique individual and corresponds one-to-one with SSNs but is assigned independently and randomly.[5] For the CPS, when a respondent did not provide an SSN or provided an SSN that could not be verified (i.e., did not match the Social Security Administration's name, gender, and birth date records for this SSN), the Census Bureau attempted to use person-associated data (such as name and address) to look up the correct SSN (then replaced by PIK) in a consolidated SSN registration file (called Census Numident). Because MSIS lacked such person-identifying fields, the Census Bureau could not perform a similar search for MSIS records lacking a validated SSN. Also, the Census Bureau could not perform a lookup for CPS sample members for whom the respondent was unwilling to provide an SSN, as this is considered by the Census Bureau as equivalent to refusal to permit SSN identification.

The linking process proceeds based on PIKs. PIKs were missing from 10 percent of the 2000 MSIS records (4.6 million out of 45.1 million total MSIS records) and 11 percent of 2001 MSIS records (5.3 million out of 48.6 million total MSIS records). In addition, PIKs were missing from 20 percent of the 2001 CPS (13 percent refused and 7 percent did not have PIKs out of the 218,269 cases in the CPS) and 20 percent of the 2002 CPS records (14 percent refused to let the Census Bureau link their data and an additional 6 percent did not have PIKs out of the 217,219 cases in the CPS). The bulk of our analyses use only those cases with PIKs in both the MSIS data and in the CPS data.

These missing PIKs in both files mean that our linked file is imperfect. To create the resulting analysis file that includes only those CPS records with SSNs representative of the full CPS survey, we reweight those CPS persons with an SSN to represent the full population covered by the CPS, including those without an SSN. Specifically, we create an "adjusted weight" by post-stratifying by age, race, sex, Hispanic ethnicity, and poverty status. For 2001, we linked 26,100 person-identified survey records, representing 36.0 million

persons on a reweighted basis, with corresponding MSIS records. For 2000, we linked 23,700 records, representing 33.5 million persons on a reweighted basis. We use this reweighted data to examine reporting errors in the CPS.[6]

RESULTS

Table 1 summarizes this part of our analysis. Row A.1 gives the unadjusted MSIS count 45.1 m/48.6 m (numbers are in millions with calendar year 2000 before the slash and calendar year 2001 after the slash). Using this number as the denominator, we derive a crude unadjusted undercount. Row B.2 gives the unadjusted CPS Medicaid count 26.1 m/27.7 m. Simple arithmetic implies an undercount of 19.0 m/20.9 m. This is the large undercount (42.1/43.0 percent) reported in the first paragraph of the paper.

Our goal is to use this linked file to refine this crude undercount. We conceptualize the CPS Medicaid undercount as having two components:

- *Imperfect Concept Alignment*: The MSIS and the CPS refer to different populations, use different definitions of Medicaid, and are not

Table 1: Counts from the Medicaid Statistical Information System (MSIS), Current Population Survey (CPS), and Linked Data Files: 2000, 2001 (Numbers in Millions)

	Calendar Year	
Selected Universe Counts	*2000*	*2001*
MSIS administrative data counts		
A.1. All people in MSIS	45.050	48.550
A.2. Minus all SCHIP-only enrollees	43.650	46.700
A.3. Minus non-full Medicaid benefit enrollees	39.750	42.200
A.4. Minus those in inst. group quarters	39.600	42.050
A.5. Minus duplicate enrollees	38.150	40.450
A.6. Minus those without PIKs (SSNs)	36.200	38.200
CPS survey counts		
B.1. All people in the CPS	277.500	279.600
B.2. Subset reported as having Medicaid	26.050	27.700
Linked data file counts		
C.1. Raw number of linked cases	0.024	0.026
C.2. Weighted number of linked cases*	33.450	36.000
C.3. Subset of linked cases reported as Medicaid*	19.090	20.550

*Weighted using the adjusted CPS person weight.
Source: 2000 and 2001 MSIS calendar year files.
SCHIP, State Children's Health Insurance Program.

comparable for additional technical reasons (e.g., CPS excludes persons residing in institutional group quarters and MSIS does not capture data on residence location, and so the only possible adjustments are based on services received such as nursing facility and intermediate care facility for the mentally retarded).

- *Response Error*: CPS responses sometimes (often) diverge from the MSIS information and these are (usually) conventional survey response errors.

Our analysis is designed to assign a magnitude to each of these undercount components and discuss the possible sources for any residual.

Imperfect Concept Alignment

Even if there was no response error (i.e., CPS responses were identical to MSIS information), the CPS counts would still diverge from the MSIS counts due to what we term "imperfect concept alignment." We define our concept of interest as unique individuals in the CPS sample frame who have full Medicaid (not partial Medicaid, nor any form of SCHIP).

Panel A of Table 1 makes adjustments to the MSIS to arrive at an estimate of the number of MSIS records we could hope to match to the CPS. From the unadjusted MSIS count (row A.1), we subtract.

- *Row A.2—SCHIP Enrollees*: States are not required to report to MSIS people in SCHIP stand-alone programs. While some states do report all of their SCHIP enrollees, to be nationally consistent we drop all SCHIP cases from our adjusted MSIS count.[7]
- *Row A.3—Partial Medicaid*: Our concept of interest is conventional health insurance and people without full Medicaid benefits are not included in our adjusted MSIS count as most of the partial Medicaid programs are not comprehensive health insurance. Partial Medicaid coverage includes programs such as family planning only or coverage only for childbirth.
- *Row A.4—Group Quarters*: The CPS sample frame only includes the noninstitutional population; thus, individuals in institutional group quarters (e.g., certified nursing facilities and prisons) and the homeless are not in the CPS sample frame and are likewise omitted from our adjusted MSIS count.
- *Row A.5—Duplicate Records*: Some people receive Medicaid in two states during a single calendar year (or have duplicate records within

a single state). Because our concept is unique individuals receiving Medicaid in a given year (that is what the CPS measures), we include only one of these records in our adjusted MSIS count.

After subtracting out the duplicate records we can compute a new denominator for a measure of the undercount that adjusts for differences in universe alignment between the CPS and the MSIS. In both 2000 and 2001 the undercount would be reduced to 32 percent in 2000 and 31 percent in 2001 (row B.2 divided by row A.5). This is an upper bound as the number of enrollees in row A.5 is probably overstated because we are unable to identify and drop records without PIKs that are duplicates of records with PIKs or other records without PIKs.

- *Row A.6—No PIK in MSIS*: These records cannot be linked to the CPS and these could be problematic cases because we are unable to verify their identity. Future work should examine whether there is a higher rate of duplication or other problem with these cases.

After removing those with no PIK in the MSIS (row A.6), the MSIS count represents unique individuals with full Medicaid benefits, who have PIKs and could be linked to the CPS universe. The counts from Table 1 are 36.2 m/38.2 m for 2000 and 2001.

Panel B presents the CPS survey estimates for those enrolled in Medicaid and the total population. Corresponding to Panel A's adjustments to the MSIS, Panel C makes adjustments to the CPS. Recall that we are trying to isolate the effect of inconsistencies aside from response error and we have a CPS file with the MSIS information appended (by matching on SSN/PIK). Thus, we can isolate the effect of inconsistencies aside from response error by tabulating the information from the reweighted CPS cases that were linked to the MSIS records. The resulting reweighted CPS count is 33.5 m/36.0 m. These CPS counts should be compared with row A.6's MSIS adjusted count of 36.2 m/38.2 m. Even if there was no measurement error in the CPS Medicaid items, there remains a CPS Medicaid undercount of 2.7 m/2.2 m. This is 7.7/5.8 percent of the corresponding universe adjusted MSIS total without the cases missing PIKs (row A.6 in Table 1). We suspect that much of the remaining mismatch results from the considerable number of individuals in the MSIS files who are not in the CPS sample frame. Our adjustment for institutional group quarters, row A.4 of Table 1, is a lower bound; that is, some of the remaining individuals are in what the Census Bureau deems institutional group quarters but are not recorded as group quarters in the MSIS data

(Roemer 2007). In addition, some of these cases may be homeless and therefore not in the CPS's dwelling unit-based sample frame.

Ongoing analyses using these linked data are exploring the magnitude of this mismatch in definitions of group quarters. Specifically, those analyses are as follows: (1) identifying long-term care residents in the MSIS and Medicaid Analytic eXtract (MAX) data (which include Medicaid utilization data), (2) identifying respondents in 2000 Census data to determine residence type (housing unit versus institutionalized group quarters), and (3) comparing addresses from Medicaid administrative data to the Census Bureau's Master Address File for the limited number of states that provided additional address information (some homeless people show a government office as their address of record). Preliminary results from those ongoing analyses suggest that they will further reduce the discrepancy between the MSIS total and the weighted number of linked CPS records by approximately 1 million, as about 1 million of the MSIS cases counted in row A.6 of Table 1 were estimated to be living in institutional group quarters (Roemer 2007).

Survey Response Errors

The previous analysis considered the extent to which the MSIS information corresponding to the linked CPS records matched the adjusted MSIS totals. We now consider the role of response errors in the CPS.[8] We begin by considering only the linked CPS records; that is, those CPS records to which we successfully appended information from MSIS. For these CPS records we have two different Medicaid counts: (1) the count implied by the reported CPS information and (2) the count implied by the appended MSIS information. Using the adjusted weights, the appended MSIS information implies that 33.5 m/36.0 m of this analysis file has Medicaid, while the CPS responses for this group imply that only 19.1 m/20.6 m of this analysis file has Medicaid. Thus, the linked survey count is 14.4 m/15.4 m (i.e., 43.0/42.7 percent) below the MSIS count that adjusts for imperfect concept alignment; that is, in both years 43 percent of those CPS respondents that the MSIS indicates have Medicaid do not report Medicaid in the CPS.[9] This large response error rate is a significant cause of the Medicaid undercount in the CPS.

Aligning concepts between the MSIS and the CPS (e.g., eliminating SCHIP enrollees, duplicates, partial benefit Medicaid cases, and institutional group quarters residents from the MSIS count) substantially reduces the size of the crude undercount from 42 and 43 percent in 2001 and 2000 to 32 percent

in 2001 and 31 percent in 2000. Nevertheless, response errors appear to be an even larger cause of the CPS Medicaid undercount. For about 41 percent of the CPS persons linked to MSIS records indicating they had Medicaid enrollment for at least 1 day in the prior calendar year, the respondent failed to report Medicaid. The aligned CPS undercount is only 32 percent in 2000 and 31 percent in 2001. Much of the difference between these two numbers (roughly 10 percent) is due to persons not linked to the MSIS but nevertheless reported in the CPS with Medicaid (e.g., SCHIP enrollees who report Medicaid in the CPS). These cases may or may not actually have Medicaid as we do not have perfect linking information. It seems likely, for example, that many people with SCHIP mistakenly report Medicaid in the survey because the programs have the same or very similar sounding names in many states. Also people may rotate between SCHIP and Medicaid when the transition is not always clear to the enrollee. Past research has demonstrated a substantial amount of Medicaid reporting among people covered by other types of public health insurance (Davern et al. 2008).

Response errors in the CPS are not random (see Table 2). Simple cross-tabulations suggest that people who were more likely to correctly report having Medicaid include those enrolled in Medicaid longer during the reference year, and those who were enrolled in both the reference year and the survey year. In addition, response errors are also related to income and age. Enrolled children were more likely to report having Medicaid than adults 18–64 years of age. Enrolled people in families with lower incomes were more likely to report Medicaid (and less likely to report some other type of coverage) and people with higher income were less likely to report Medicaid (but were more likely to report some other type of coverage). Finally, those with imputed health insurance values are much less likely to be correct than those with edited or reported values.[10]

There are also some people for whom we did not append MSIS information who do report Medicaid. Some, but not all, of these cases are "response errors." Table 3 shows the complement to Table 2 and includes the reweighted CPS estimates for those cases that were linkable (i.e., had an SSN) but did not link to the MSIS. We are most interested in the reports of "Medicaid only" and "Medicaid plus some other type of coverage," as the Medicaid report could be in error. In Table 2 we are relatively confident that the combination of a CPS report of no Medicaid and an MSIS link indication of Medicaid is a response error. We are less confident that the converse—that is, a CPS report of Medicaid and no MSIS report of Medicaid—is a response error. This pattern could be due to a missing MSIS SSN causing us to fail to

Table 2: Current Population Survey (CPS) Responses to the Health Insurance Coverage Items by People Linked to Medicaid Statistical Information System (MSIS) by Selected Characteristics: Survey Reference Year of 2001 (2002 CPS Survey Year)

Selected Characteristics	*Persons Coded Medicaid Only (%)*	*Standard Error (%)*	*Persons Coded Medicaid and Something Else (%)*	*Standard Error (%)*	*Persons Coded with Some Other Type of Health Insurance Coverage (%)*	*Standard Error (%)*	*Persons Coded as Being Uninsured (%)*	*Standard Error (%)*	*Total (in Thousands) (%)*
Age									
0–5	50.6	0.9	10.9	0.6	24.5	0.8	13.7	0.6	7,740
6–14	47.8	0.9	10.8	0.5	25.1	0.7	16.3	0.6	9,080
15–17	44.1	1.8	9.8	1.1	25.5	1.6	19.6	1.4	2,040
18–44	38.1	0.8	11.7	0.5	24.1	0.7	26.1	0.7	10,800
45–64	40.9	1.3	25.0	1.2	21.6	1.1	12.5	0.9	3,520
65 and older	0.7	0.3	58.6	1.5	39.3	1.5	1.4	0.4	2,800
Race/ethnicity									
White	40.3	0.5	17.1	0.4	25.7	0.5	17.1	0.4	23,450
Black	42.8	0.8	14.1	0.6	25.1	0.7	17.8	0.6	10,100
Native American	46.3	2.5	11.1	1.6	22.2	2.1	20.4	2.0	1,080
Asian Pacific Islander	35.7	2.1	17.1	1.6	30.0	2.0	17.1	1.6	1,400
Sex									
Male	43.0	0.7	15.1	0.5	25.7	0.6	16.4	0.5	14,550
Female	39.6	0.5	16.8	0.4	25.5	0.5	18.1	0.4	21,400
Hispanic ethnicity									
Hispanic	44.7	0.9	11.6	0.6	21.2	0.8	22.5	0.8	7,740
Non-Hispanic	40.0	0.5	17.3	0.4	26.8	0.4	15.9	0.4	28,250
Poverty level									

continued

Table 2. *Continued*

Selected Characteristics	*Persons Coded Medicaid Only (%)*	*Standard Error (%)*	*Persons Coded Medicaid and Something Else (%)*	*Standard Error (%)*	*Persons Coded with Some Other Type of Health Insurance Coverage (%)*	*Standard Error (%)*	*Persons Coded as Being Uninsured (%)*	*Standard Error (%)*	*Total (in Thousands) (%)*
0–49%	57.9	1.0	7.9	0.5	11.8	0.6	22.1	0.8	6,600
50–74%	55.2	1.2	14.9	0.9	14.0	0.9	15.8	0.9	4,420
75–99%	44.0	1.2	23.7	1.0	19.5	0.9	13.3	0.8	4,820
100–124%	43.0	1.3	20.3	1.0	21.7	1.0	15.5	0.9	4,140
125–149%	38.0	1.4	15.2	1.0	30.4	1.3	17.7	1.1	3,160
150–174%	34.0	1.5	16.3	1.1	32.6	1.4	16.3	1.1	2,820
175–199%	31.1	1.7	15.5	1.3	35.9	1.7	18.4	1.4	2,060
200% plus	22.5	0.8	17.5	0.7	42.0	0.9	18.0	0.7	8,000
Enrolled in survey year and length of time enrolled in reference year									
Eligible for <61 days of year	25.6	2.4	11.6	1.8	32.6	2.6	27.9	2.5	860
Eligible for 61–180 days of year	31.7	1.4	12.4	1.0	31.7	1.4	24.1	1.3	2,900
Eligible for >180 days of year	47.0	0.5	18.4	0.4	21.1	0.4	13.4	0.3	26,150
Not enrolled in survey year and length of time enrolled in reference year									
Eligible for <61 days of year	11.9	1.8	4.8	1.2	50.0	2.8	31.0	2.6	840

continued

Eligible for 61–180 days of year	14.0	1.3	6.5	0.9	44.1	1.9	35.5	1.8	1,860
Eligible for >180 days of year	26.3	1.2	10.5	0.9	35.1	1.3	26.9	1.2	3,420
Imputed or edited or reported									
Edited	51.3	2.1	48.7	2.1	0.0	0.0	0.0	0.0	1,560
Imputed	19.5	0.9	15.0	0.8	40.4	1.1	24.7	1.0	5,340
Reported	44.3	0.5	14.6	0.3	24.1%	0.4	16.9	0.4	29,100
Overall	41.0	0.4	16.1	0.3	25.6	0.4	17.4	0.3	36,000
Total unweighted count	10,400		3,800		6,200		3,500		23,900

Standard errors computed using the generalized variance estimate approach (see Davern et al. 2006, 2007b for more information on limitations associated with this approach). Owing to rounding, total values may not equal column or row sums exactly.

Table 3: Current Population Survey (CPS) Responses to the Health Insurance Coverage Items by Not Linked to Medicaid Statistical Information System (MSIS) by Selected Characteristics: Survey Reference Year of 2001 (2002 CPS Survey Year)

Selected Characteristics	*Persons Coded Medicaid Only (%)*	*Standard Error (%)*	*Persons Coded Medicaid and Something Else (%)*	*Standard Error (%)*	*Persons Coded with Some Other Type of Health Insurance Coverage (%)*	*Standard Error (%)*	*Persons Coded as Being Uninsured (%)*	*Standard Error (%)*	*Total (in Thousands) (%)*
Age									
0–5	3.5	0.2	1.8	0.2	86.5	0.4	8.5	0.4	15,800
6–14	2.2	0.1	1.9	0.1	85.3	0.3	10.5	0.3	28,850
15–17	1.4	0.2	1.6	0.2	85.1	0.6	11.9	0.6	8,720
18–44	1.0	0.1	0.7	0.0	77.6	0.2	20.6	0.2	98,750
45–64	0.8	0.1	1.0	0.1	84.6	0.2	13.6	0.2	60,950
65 and older	0.1	0.0	3.9	0.2	95.2	0.2	0.7	0.1	29,850
Race/ethnicity									
White	1.0	0.0	1.3	0.0	85.0	0.1	12.8	0.1	203,400
Black	2.9	0.2	2.7	0.2	73.1	0.4	21.2	0.4	26,650
Native American	2.4	0.5	1.6	0.4	64.6	1.5	30.7	1.5	2,540
Asian Pacific Islander	1.4	0.2	1.1	0.2	78.7	0.6	18.6	0.6	11,050
Sex									
Male	1.2	0.1	1.3	0.1	82.1	0.2	15.5	0.2	121,850
Female	1.2	0.1	1.5	0.1	84.3	0.2	12.9	0.2	121,750
Hispanic ethnicity									
Hispanic	3.0	0.2	1.5	0.1	61.7	0.5	33.9	0.5	24,400
Non-Hispanic	1.0	0.0	1.4	0.0	85.6	0.1	12.0	0.1	219,200
Poverty level									

continued

0–49%	7.0	0.5	2.1	0.3	44.3	1.0	46.6	1.0	6,820
50–74%	8.4	0.7	3.4	0.5	45.3	1.3	42.9	1.3	4,060
75–99%	4.9	0.4	3.3	0.4	52.3	1.0	39.5	1.0	6,120
100–124%	4.3	0.4	2.8	0.3	59.8	0.9	33.2	0.9	7,820
125–149%	3.0	0.3	2.8	0.3	68.1	0.8	25.8	0.7	9,840
150–174%	1.9	0.2	2.3	0.2	72.0	0.7	23.7	0.7	10,300
175–199%	1.8	0.2	1.7	0.2	74.7	0.7	21.5	0.6	10,900
200% plus	0.4	0.0	1.1	0.0	89.3	0.1	9.2	0.1	187,800
Imputed or edited or reported									
Edited	42.4	2.3	55.9	2.4	0.0	0.0	0.0	0.0	1,180
Imputed	2.9	0.2	5.1	0.2	71.0	0.4	21.2	0.4	31,550
Reported	0.7	0.0	0.6	0.0	85.5	0.1	13.2	0.1	210,850
Overall	1.2	0.0	1.4	0.0	83.2	0.1	14.2	0.1	243,600
Total unweighted count	1,850		1,950		126,500		16,800		147,000

Standard errors computed using the generalized variance estimate approach (see Davern et al. 2006, 2007b for more information on limitations associated with this approach). Owing to rounding, total values may not equal column or row sums exactly.

link the MSIS record to the CPS. Again, this pattern would appear to be a response error when it was not. In addition, someone who had SCHIP in a state where SCHIP and Medicaid share the same program name (e.g., Hoosier Healthwise in Indiana can be Medicaid or SCHIP) may say yes to the Medicaid question that includes the same program name as the SCHIP program they are enrolled in. So their answer is "correct" in that the Medicaid question asks about the same program name as SCHIP, but we failed to match them to Medicaid cases included in the MSIS.

Finally, other patterns would induce errors in the total count of the uninsured. For example, a person could have answered "yes" to Medicaid but have only partial Medicaid coverage that we do not consider a full-benefits Medicaid health insurance coverage plan, and the person should therefore be counted as "uninsured." Alternatively, a person may answer having Medicaid only but not ever have been enrolled in Medicaid nor have another type of insurance coverage. These response errors would lead to an underestimate of the number of uninsured.

Table 3 shows that 2.6 percent of the CPS cases that were not linked reported having Medicaid. Interestingly, 1.2 percent of the unlinked CPS cases reported "Medicaid only" as their coverage and would have otherwise been uninsured, while the other 1.4 percent reported having some other health insurance. In this table, simple tabulations suggest that being younger, Hispanic, and having a lower poverty level to income threshold is associated with higher levels of reporting Medicaid but not linking to the MSIS. Also, if the person's data were edited, s/he is more likely to say Medicaid in the CPS and not link to the MSIS. This likely happens because the CPS edits cases to have Medicaid if someone in the household reports Supplemental Security Income and no one is ever edited to "not have Medicaid." Thus, edited cases are always coded to have Medicaid in the CPS and cannot be uninsured.

Call et al. (2008) reviewed existing estimates of Medicaid response error rates in a variety of state surveys and found much lower rates. Like Klerman, Ringel, and Roth's (2005) earlier work on the CPS in California, our error rates for national CPS data are much higher than the estimates presented by Call et al. (2008) for state surveys. Three factors might plausibly explain the lower response accuracy in the CPS:

1. The long reference period (up to 16 months for coverage; that is, asking in February through April about health insurance coverage over the entire January to December period of the previous year).[11]
2. The household-level screening included in the health insurance survey items (e.g., Does anyone in the household have Medicaid?)

rather than person-level survey items (e.g., Does Jim have Medicaid?).

3. Differential methods employed to examine the problem. Most of the studies reported on by Call et al. (2008) used a list frame of enrollees in a single state and attempted to contact them over the telephone to ask them a health insurance screener. The approach used in our CPS analysis was to link the CPS records (after the fact) to the Medicaid enrollment data. It is possible that these significantly varying methodologies could lead to different rates of response error because of selection bias into the analytic sample.

Earlier literature provides support for the first two explanations (Hess et al. 2001; Call et al. 2008) and more research should be conducted to determine the potential impact of the third explanation.

DISCUSSION

Our analysis has considered two major causes of the Medicaid undercount: imperfect concept alignment between the MSIS and the CPS and CPS response errors. In this section, we review our findings on each of these issues. We also examine the implications of our findings for using the CPS to estimate the number of people who lacked health insurance coverage for all of the preceding calendar year.

Accounting for the Undercount

Aligning concepts between the MSIS and the CPS (e.g., eliminating SCHIP enrollees, duplicates, partial benefit Medicaid cases, and institutional group quarters residents from the MSIS count) reduces the size of the crude undercount from 42 and 43 percent in 2001 and 2000 to 32 percent in 2000 and 31 percent in 2001. Both of these estimates make the assumption that the MSIS records without SSN and resulting PIK are all unique individuals and are not duplicates with other MSIS records (with or without PIKs). This assumption should be explored further as the cases with missing identifying information may have a higher rate of duplication of invalid records than the unduplicated cases. Assuming that none of the records without PIKs are duplicates of records with PIKs or of each other will greatly inflate the estimate of the undercount. Future work into the nature of these cases missing identifying information on MSIS is crucial to a better understanding of the problem.

Response errors appear to be an even larger cause of the CPS Medicaid undercount. About 43 percent of the CPS persons linked to MSIS records indicating they had Medicaid enrollment for at least 1 day in the prior calendar year failed to report having Medicaid. In the total CPS Medicaid count, however, this very high error rate is partially offset by errors in the other direction; that is, people who report in the CPS that they have Medicaid but cannot be linked to the MSIS.

Implications for Estimates of the Uninsured

Our analysis has focused on the CPS Medicaid undercount. While this CPS Medicaid undercount is of considerable interest to survey researchers, policy interest in the CPS Medicaid undercount is due primarily to the implications of the undercount for the count of the number of uninsured people in the United States.

Translating our results on the CPS Medicaid undercount into adjustments to the count of the uninsured—even for the years for which we have linked data—is not straightforward. In order to affect the count of the uninsured a respondent must misreport Medicaid and not report other health insurance. While 43 percent of the Medicaid cases we linked failed to report having Medicaid, many of them did report some other source of health insurance coverage. For calendar year 2001, the number of cases in the CPS that were coded as uninsured but had at least 1 day of MSIS eligibility during the calendar year was 6 million people (in Table 2, 17.4 percent of the weighted matched cases were coded as being uninsured and the weighted matched total was 36,000,000). This 6 million figure is a rough approximation (in which we report whole millions of people to emphasize the roughness of the approximation) of the downward adjustment to the CPS estimate of the uninsured. It accounts for those who report no health insurance, but our linked data suggest that they have Medicaid.

In addition to this adjustment for people with Medicaid coverage who report no health insurance, our linked data suggest an offsetting adjustment for people who only report Medicaid but who, our linked data suggest, do not have Medicaid. Our linked data suggest that 3 million people report Medicaid and no other health insurance, but we did not link them to the MSIS (in Table 3, 1.2 percent of the unlinked CPS cases reported having Medicaid only and were not matched and the total weighted population was 243,600,000, of which 1.4 million came from edited or imputed cases). For some of these people, their MSIS Medicaid records could be lacking PIKs (SSNs), so we have treated them (potentially incorrectly) as not having Medicaid. It seems likely that others—but it is unclear how many—have confused Medicaid with

some other health insurance such as SCHIP. For those individuals, their reporting error (i.e., a CPS report of Medicaid but no Medicaid in MSIS) would have no effect on the number of uninsured individuals. Finally, we know that some of the partial Medicaid enrollees reported their Medicaid "coverage" as Medicaid even though it is not a comprehensive insurance program. In our ongoing analysis, we will continue to investigate the impact of people who report having Medicaid in the CPS but do not have SSNs and therefore cannot be linked to the MSIS enrollment data. An adjusted estimate of the number of uninsured will require refining both of these adjustments.

Our current study implies that the originally published CPS estimate of 41.2 million people who lacked coverage for all of 2001 (Mills 2002) should be lowered by between 3 and 6 million people, with a more precise answer to be determined through future modeling. This correction would be in addition to other downward revisions to the number of uninsured in the published CPS estimates that have been demonstrated due to editing (1.4 million in 2001) and imputation (2.5 million in 2001) problems in the CPS (Davern et al. 2007b; Lee and Stern 2007). The editing issue noted by Lee and Stern (2007) has been incorporated into the current CPS official estimates and imputation corrections suggested by Davern et al. (2007b) are being tested by the Census Bureau for integration into the CPS estimates. Corrections for the Medicaid reporting errors are unlikely to make it into the official estimates given the lag needed to verify the survey data; however, researchers can use methods like the one suggested in Davern, Klerman, and Ziegenfuss (2007a) to improve the CPS microdata for their own analysis.

CONCLUSIONS

This paper attempts a basic accounting for the crude CPS and MSIS undercount. We note that the undercount is caused by a combination of inconsistencies between the MSIS and CPS universes and survey response errors. While an exact accounting is not possible at this point, we demonstrate that both issues are substantively important. We reach these conclusions using a new dataset created by linking MSIS files to individual CPS records for calendar years 2000 and 2001. We find that reporting errors are very common, with 43 percent of the CPS cases that the MSIS implies having Medicaid not coded to have Medicaid in the CPS. We also find that comparing the raw MSIS to the survey counts is inappropriate, as the raw MSIS count includes many people not eligible to be in the survey (e.g., institutional group quarters

residents) and counts some people twice (people who received Medicaid in two or more states are legitimately included in the MSIS count more than once) and that many people in the MSIS count are missing key identifying information that allows the record to be verified for linking purposes.

Our findings have direct implications for research that is done using the CPS survey data for important policy analysis of the Medicaid and the uninsured populations. Many people who actually have Medicaid, according to administrative data, fail to report it and this should be considered when estimating take-up rates for public programs using survey data. Also, when examining the size of the Medicaid-eligible but uninsured population, some consideration should be given to the fact that many people coded in the CPS as being uninsured are shown to have had Medicaid. Using the results presented in Tables 1–3, it is possible to estimate the size of some of these adjustments needed to properly use the survey data for policy research purposes.

ACKNOWLEDGMENTS

Joint Acknowledgments/Disclosure Statement: This paper was made possible by grant no. 052084 from the Robert Wood Johnson Foundation to the State Health Access Data Assistance Center (Michael Davern, PI) with additional support supplied by the Office of the Assistant Secretary for Planning and Evaluation (ASPE), the Centers for Medicare and Medicare Services (CMS), the National Center for Health Statistics (NCHS), and the U.S. Census Bureau. This paper has undergone a limited review by all the participating organizations in accordance with existing agreements among these organizations. The views expressed are those of the authors and do not represent official positions of ASPE, NCHS, CMS, the U.S. Census Bureau, Abt Associates, or Rand Corporation. For this paper we would also like to thank Bill Clark and Karen Soderberg.

NOTES

1. Based on raw unadjusted CPS estimates of the number of people with Medicaid and MSIS administrative data counts. In calendar year 2001 (using the 2002 CPS) the MSIS count was 48.6 million and the CPS count was 27.7 million. In calendar year 2000 (CPS year 2001) the MSIS count was 45.1 million and the CPS count was 26.1 million. See Table 1 for details.
2. We used the 2001 expanded CPS sample file; see Davern et al. (2003) for a description.
3. Respondents with Indian Health Service coverage only were not considered insured.

4. Despite the fact that some persons were enrolled in a Medicaid eligibility group identified as having restricted Medicaid benefits for pregnancy services, analysis of MSIS claims data suggests that people enrolled in this group use many types of Medicaid services.
5. See the SNACC Phase I report for details (State Health Access Data Assistance Center 2007).
6. For more details on the reweighting process and for an analysis of the missing PIKs on the CPS, please see the SNACC Phase II report (State Health Access Data Assistance Center 2008).
7. The number of children ever enrolled in SCHIP in fiscal year 2001 was 4.6 million (CMS 2006).
8. In this analysis we only report data in Tables 2 and 3 from calendar year 2001 to save space. The calendar year 2000 results are very similar and are available from the corresponding author on request.
9. We use "report" to include imputed and allocated responses as well as those that were reported by a household member. For an in-depth analysis of the complexity of the household and the impact of the relationship to reference person on reporting error, please see Pascale, Roemer, and Resnick (2007).
10. This is expected, as the goal of imputation is to be correct in the aggregate population estimates and not to obtain any individual person's health insurance status correct.
11. Other papers expressing concern about the long CPS recall period include Congressional Budget Office (2003), Lewis, Elwood, and Czajka (1998), and Swartz (1986).

REFERENCES

Blewett, L. A., and M. Davern. 2006. "Meeting the Need for State Level Estimates of Health Insurance Coverage: What Has Been Done and How It Can Be Improved." *Health Services Research* 41 (3): 946–75.

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.

Call, K. T., G. Davidson, M. Davern, L. A. Blewett, and R. Nyman. 2008. "The Medicaid Undercount and Bias to Estimates of Uninsurance: New Estimates and Existing Evidence." *Health Services Research* 42 (3): 901–14.

CMS. 2006. "FY 2001 Annual Enrollment Reports." National SCHIP Policy, SCHIP Enrollment Reports web page. Washington, DC: Department of Health and Human Services [accessed on August 18, 2008]. Available at http://www.cms.hhs.gov/NationalSCHIPPolicy/SCHIPER/itemdetail.asp?filterType=none&filterByDID=-99&sortByDID=2&sortOrder=ascending&itemID=CMS056905&intNumPerPage=10

———. 2007a. "Medicaid Data Sources—General Information MSIS Data web page." Baltimore, MD: Centers for Medicare and Medicaid Services [accessed on August 1, 2008]. Available at http://www.cms.hhs.gov/medicaidDataSourcesGenInfo/02_MSISData.asp

———. 2007b. "Medicaid Statistical Information Statistics (MSIS) Overview web page." Baltimore, MD: Centers for Medicare and Medicaid Services [accessed on February 6, 2008]. Available at http://www.cms.hhs.gov/MSIS/

Congressional Budget Office. 2003. *How Many People Lack Health Insurance and for How Long?* Economic and Budget Issue Brief. Washington, DC: Congressional Budget Office.

Davern, M., L. A. Blewett, B. Bershadsky, K. T. Call, and T. Rockwood. 2003. "State Variation in SCHIP Allocations: How Much Is There, What Are Its Sources, and Can It Be Reduced?" *Inquiry* 40 (2): 184–97.

Davern, M., K. T. Call, J. Ziegenfuss, G. Davidson, T. J. Beebe, and L. A. Blewett. 2008. "Validating Health Insurance Coverage Survey Estimates: A Comparison between Self-Reported Coverage and Administrative Data Records." *Public Opinion Quarterly* 72 (2): 241–59.

Davern, M., A. Jones Jr., J. LepKowski, G. Davidson, and L. A. Blewett. 2006. "Unstable Inferences? An Examination of Complex Survey Sample Design Adjustments Using the Current Population Survey for Health Services Research." *Inquiry* 43 (3): 283–97.

Davern, M., J. A. Klerman, and J. Ziegenfuss. 2007a. "Medicaid Under-Reporting in the Current Population Survey and One Approach for a Partial Correction." Minneapolis, MN: State Health Access Data Assistance Center [accessed on December 26, 2008]. Available at http://www.shadac.org/publication/medicaid-under-reporting-in-cps-and-one-approach-partial-correction

Davern, M., H. Rodin, L. A. Blewett, and K. T. Call. 2007b. "Are the CPS Uninsurance Estimates Too High? An Examination of Imputation." *Health Services Research* 42 (5): 2038–55.

Glied, S., D. Remler, and J. G. Zivin. 2002. "Inside the Sausage Factory: Improving Estimates of the Effects of Health Insurance Expansion Proposals." *Milbank Quarterly* 80 (4): 603–35.

Hess, J., J. Moor, J. Pascale, J. Rothgeb, and C. Keeley. 2001. "The Effects of Person-Level Versus Household-Level Questionnaire Design on Survey Estimates and Data Quality." *Public Opinion Quarterly* 65 (4): 574–84.

Klerman, J. A., J. S. Ringel, and B. Roth. 2005. *Under-Reporting of Medicaid and Welfare in the Current Population Survey.* Santa Monica, CA: RAND.

Lee, C. H., and S. M. Stern. 2007. "Health Insurance Estimates from the U.S. Census Bureau: Background for a New Historical Series." Washington, DC: U.S. Census Bureau [accessed on December 26, 2008]. Available at http://www.census.gov/hhes/www/hlthins/usernote/revhlth_paper.pdf

Lewis, K., M. Elwood, and J. Czajka. 1998. *Counting the Uninsured: A Review of the Literature.* Washington, DC: The Urban Institute.

Lynch, V., D. Resnick, J. Staveley, and C. Taeuber. 2007. "Differences in Estimates of Public Assistance Recipiency between Surveys and Administrative Records." Paper presented at the meeting of the Federal Committee on Statistical Methodology, Arlington, VA, November 5–7, 2007 [accessed on December 26, 2008]. Available at http://www.fcsm.gov/events/papers2007.html

Mills, R. J. 2002. "Health Insurance Coverage: 2001." Current Population Reports No. P60-220. Washington, DC: U.S. Census Bureau. Available at http://www.census.gov/prod/2002pubs/p60-220.pdf

Pascale, J., M. Roemer, and D. Resnick. 2007. "Medicaid Underreporting in the CPS: Results from a Record Check Study." Research Report Series, Survey Methodology #2007-40. Suitland, MD: U.S. Census Bureau [accessed August 18, 2008]. Available at http://www.census.gov/srd/papers/pdf/rsm2007-40.pdf

Roemer, M. 2007. *Phase III: MSIS People Living in Institutions.* Presentation at the project conference, An Examination of the Medicaid Undercount in the Current Population Survey (CPS): Results and Policy Implications, Washington, DC, July 17.

State Health Access Data Assistance Center. 2007. *Phase I Research Results: Overview of National Medicare and Medicaid Files.* Research Project to Understand the Medicaid Undercount: The University of Minnesota's State Health Access Data Assistance Center, the Centers for Medicare and Medicaid Services, the Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation, and the U.S. Census Bureau. Minneapolis, MN: University of Minnesota.

———. 2008. *Phase II Research Results: Examining Discrepancies between the National Medicaid Statistical Information System (MSIS) and the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC).* Research Project to Understand the Medicaid Undercount: The University of Minnesota's State Health Access Data Assistance Center, the Centers for Medicare and Medicaid Services, the Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation, and the U.S. Census Bureau. Minneapolis, MN: University of Minnesota.

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. 2002a. *Current Population Survey Technical Paper: Design and Methodology.* Report #TP63RV. Washington, DC: U.S. Census Bureau.

———. 2002b. *Current Population Survey Technical Documentation. March 2001 Annual Demographic File, G-3.* Washington, DC: U.S. Census Bureau.

———. 2003. *Current Population Survey Technical Documentation, March 2002 Annual Demographic File, G-3.* Washington, DC: U.S. Census Bureau.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.