

© Health Research and Educational Trust
DOI: 10.1111/1475-6773.12308
RESEARCH ARTICLE

Measurement Error in Public Health Insurance Reporting in the American Community Survey: Evidence from Record Linkage

Michel H. Boudreaux, Kathleen Thiede Call, Joanna Turner, Brett Fried, and Brett O'Hara

Objective. Examine measurement error to public health insurance in the American Community Survey (ACS).

Data Sources/Study Setting. The ACS and the Medicaid Statistical Information System (MSIS).

Study Design. We tabulated the two data sources separately and then merged the data and examined health insurance reports among ACS cases known to be enrolled in Medicaid or expansion Children's Health Insurance Program (CHIP) benefits.

Data Collection/Extraction Methods. The two data sources were merged using protected identification keys. ACS respondents were considered enrolled if they had full benefit Medicaid or expansion CHIP coverage on the date of interview.

Principal Findings. On an aggregated basis, the ACS overcounts the MSIS. After merging the data, we estimate a false-negative rate in the 2009 ACS of 21.6 percent. The false-negative rate varies across states, demographic groups, and year. Of known Medicaid and expansion CHIP enrollees, 12.5 percent were coded to some other coverage and 9.1 percent were coded as uninsured.

Conclusions. The false-negative rate in the ACS is on par with other federal surveys. However, unlike other surveys, the ACS overcounts the MSIS on an aggregated basis. Future work is needed to disentangle the causes of the ACS overcount.

Key Words. Medicaid, CHIP, survey methods, American Community Survey

The "Medicaid undercount" refers to the tendency of population surveys to undercount Medicaid enrollment relative to administrative records (Call et al. 2012). The undercount is a function of definitional differences between survey and administrative sources and false-negative reporting errors among respondents (Davern et al. 2009b). False-negative errors occur when a person known to have Medicaid is coded as not having that coverage. The false-negative rate

in the 2005 Current Population Survey (CPS) was 42.9 percent (University of Minnesota's State Health Access Data Assistance Center, SNACC Phase V). In the 2002 National Health Interview Survey (NHIS) the false-negative rate was 32.2 percent, and in the 2003 Medical Expenditure Panel Survey it was 17.5 percent (University of Minnesota's State Health Access Data Assistance Center SNACC Phase IV, SNACC Phase VI). In addition to biasing estimates of Medicaid, undercounted Medicaid enrollees that are not classified into other coverage types upwardly bias counts of the uninsured. Despite their limitations, survey data play an important role in research and policy making because they are timelier than administrative records and are the only source of information on the uninsured (Blewett and Davern 2006). Documenting measurement error provides researchers a gauge for the bias in a particular survey and a basis to compare the quality of alternative data.

This article examines measurement error in Medicaid and other public health insurance in the 2009 American Community Survey (ACS). The ACS is distinct among federal surveys measuring health insurance. Its sample size of over 4 million persons makes it the only federal survey capable of producing statistically reliable single-year health insurance estimates for the entire age distribution at the state and substate level (Davern et al. 2009a). The ACS differs from other federal surveys in its measurement of health insurance. It relies heavily (but not exclusively) on a mailed questionnaire that uses a single question stem with yes/no responses for six comprehensive insurance types. It combines Medicaid, CHIP, and other means-tested public coverage into a single response option, and it lacks state-specific Medicaid names or a specific mention of CHIP and other state-based programs.

To examine measurement error in the ACS, we use a version of the survey that has been linked with enrollment records from the Medicaid Statistical Information System (MSIS). The MSIS is not a perfect gold standard. It is compiled by aggregating files submitted by the states. In some states, the scope of benefits or enrollment period is reported with error (CMS 2013a). Nonetheless,

Address correspondence to Michel H. Boudreaux, Ph.D., Department of Health Services Administration, School of Public Health, University of Maryland, College Park, 3310 School of Public Health Building, College Park, MD 20742; e-mail: mhb@umd.edu. Kathleen Thiede Call, Ph.D., is with the Division of Health Policy and Management and State Health Access Data Assistance Center, School of Public Health, University of Minnesota, Minneapolis, MN. Joanna Turner, M.S., and Brett Fried, M.S., are with the State Health Access Data Assistance Center, University of Minnesota, Minneapolis, MN. Brett O'Hara, Ph.D., is with the Social, Economic, and Housing Statistics Division, U.S. Census Bureau, Suitland, MD.

the MSIS is the most useful tool for assessing the quality of Medicaid data collected by the ACS.

Concept Alignment and Measurement Error

Discrepancies between administrative counts and survey estimates arise from divergent concept alignment and survey error (Davern et al. 2009b). Divergent concept alignment refers to the fact that administrative databases and surveys track different populations and use different definitions of health insurance. In addition, when the MSIS is treated as a referent dataset in a validation study, divergent concept alignment encompasses error contained in the MSIS. The ACS sample represents the entire population (housing units and group quarters) and, in that sense, the ACS is the best aligned survey to the MSIS.

The MSIS and ACS, however, have different definitions of health insurance. The MSIS tracks enrollment in full and partial benefit plans that are financed by Medicaid and a subset of Children's Health Insurance Programs (CHIP). In contrast, the ACS uses a single response item intended to measure full-benefit health insurance provided through Medicaid, all CHIP plans, and state-specific plans (the response item reads "Medicaid, Medical Assistance, or any kind of government-assistance plan for those with low incomes or a disability").¹ The restriction of the ACS to comprehensive coverage contributes to undercounting enrollment in the MSIS, even in the absence of survey error. The inclusion of all public programs in a single measure contributes to overcounting in the MSIS.

After accounting for concept alignment, remaining differences can be attributed to survey error. Survey error can arise from several sources. Most of the literature focuses on response error where the reference period plays an important role. The CPS asks respondents to recall coverage held up to 16 months prior to the survey, and it tends to underestimate Medicaid enrollment to a stronger degree than surveys that have a point-in-time question (Klerman et al. 2009; Call et al. 2012). The ACS fields a point-in-time measure which ascertains coverage status for the date of interview and avoids recall error. However, other factors contribute to misreporting: the structure of response options, household composition and the flow of the instrument (person or household centered), the size and scope of the questionnaire, and respondent understanding of health insurance (Pascale 2008). Survey error can also be introduced or corrected by data processing (Lynch, Boudreaux, and Davern 2009). In this article, we examine variation in survey error across

versions of the data that have undergone different stages of data processing (explicit reports, imputation, and logical edits).

DATA AND METHODS

The American Community Survey

The ACS is conducted by the U.S. Census Bureau and has a weighted response rate of 98 percent (U.S. Census Bureau 2014). Data are collected via mail-out/mail-back questionnaires (68 percent of unweighted housing unit responses), with telephone (13 percent), and in-person (18 percent) nonresponse follow-up.² The 2009 file contains 4.5 million individual records. Our primary focus is the civilian noninstitutional population. This aligns our analysis with the universe that is frequently used in health policy research and the universe for which the Census Bureau publishes health insurance estimates.³ The size of the civilian noninstitutional sample is 4.4 million.

The questionnaire intends to capture six comprehensive health insurance types: employer-sponsored insurance, coverage purchased directly from an insurer, Medicare, Medicaid or other means-tested coverage, TRICARE, and coverage from the Department of Veterans Affairs. Uninsurance is a residual category that is defined when a respondent is not classified into any coverage type (Figure S1).⁴ We refer to the means-tested coverage category as “Medicaid Plus.”⁵ The broader Medicaid Plus definition in the ACS compared to the enrollment information in the MSIS restricts us to estimating one component of error. We only observe false-negatives and do not observe false-positives. If a record is found on the MSIS, the survey respondent should be coded to Medicaid Plus. However, if a record is not on the MSIS, it is ambiguous whether a Medicaid Plus response is an error or a correct report of another valid coverage type.⁶

Despite the relative simplicity of the ACS, its estimates of public, private, and any insurance coverage track closely with other federal surveys (Turner, Boudreaux, and Lynch 2009; Boudreaux et al. 2011). The exception is directly purchased coverage. All surveys overstate the size of the individual market, but the problem is extreme in the ACS (Mach and O’Hara 2011; Abraham, Karaca-Mandic, and Boudreaux 2013). Previous work using similar data to those presented here found that Medicaid misreporting plays a minimal role in the survey’s direct purchase estimate (Boudreaux et al. 2014).

The Medicaid Statistical Information System

The MSIS is compiled by the Centers for Medicare and Medicaid Services (CMS) from quarterly updates provided by the states (CMS 2013b). The MSIS file we use for linking consists of an eligibility record for individuals enrolled in full-benefit Medicaid or expansion CHIP. The states have the option to use their CHIP funds to either expand Medicaid (“expansion CHIP”), to set up a separate program (“stand-alone CHIP”), or to do some combination of the two. In 2009, 18 states had only stand-alone CHIP, 6 states and DC had only expansion CHIP, and 26 had a combination.⁷ Overall, 28 percent of CHIP enrollees were enrolled in Medicaid expansion programs and are included in our data (CMS 2012). We did not include stand-alone CHIP programs because those data are not consistently reported by the states (Czajka 2012; Klerman, Plotzke, and Davern 2012). The exclusion of stand-alone CHIP limits our generalizability to Medicaid and expansion CHIP.

We primarily used MSIS data pertaining to calendar year 2009. We cleaned the MSIS by removing deceased persons (if a person was coded as deceased in MSIS but had an ACS interview they were retained) and duplicated persons (2.9 million or 5.7 percent of the original file). In a secondary analysis, we also use the 2008 MSIS to contrast the false-negative rate in the 2008 and 2009 ACS.

Data Linkage, Weighting, and Defining Known Enrollees

We linked the ACS and MSIS using protected identification keys (PIKs), which are common to both data. PIKs are anonymous individual identifiers that are derived from personally identifiable information (Wagner and Layne 2012). After PIKs were added to each file, all personally identifiable information was removed. A total of 412,000 cases (9.2 percent) in the ACS lack a valid PIK, and the probability of missing a PIK is not random. For example, in the sample lacking a valid PIK, the weighted percentage of Hispanics was 34.4 percent, but in the valid PIK sample it was 13.8 percent (Table S1). The majority of missing PIKs on the ACS are due to incomplete matching variables on the ACS (e.g., name) rather than coverage error in the underlying PIK database which is built from the Social Security Administration’s Numident file (Mulrow et al. 2011). This suggests that the missing PIKs are not primarily composed of respondents that lack social security numbers.

We drop all ACS cases that do not have a valid PIK, and to ensure that the subset of ACS cases that have PIKs represent the full ACS file, we

reweighted the data by poststratifying the sample weights by state of residence, age, race/ethnicity, poverty status, and health insurance status. Table S1 provides detailed results describing characteristics of the sample before and after reweighting. The bias to the estimate of the uninsured was -2.1 percentage points prior to and 0.0 after poststratification.⁸

In addition to missing PIKs in the ACS, 4.4 million records (6.0 percent) in the MSIS lack a valid merging key and are dropped from the linked analysis. Missing PIKs in the MSIS are caused by missing or invalid social security numbers (Czajka and Verghese 2013). The majority of missing PIKs (68 percent) are partial benefit enrollees that are not eligible for linking. Of full-benefit enrollees, most with missing PIKs are under 5 years old (Czajka and Verghese 2013). Missing PIKs in the MSIS could cause bias in our estimate of the false-negative rate because a portion of enrollees observed in the MSIS cannot be merged and they could have different reporting patterns. We have no solution to this problem and it remains a limitation.

The MSIS tracks days of enrollment per month, and the ACS tracks coverage held on the day of interview. We used an algorithm (see the technical appendix) to ascertain if a case was enrolled on the day of interview. We considered only full-benefit coverage from the MSIS (including pregnancy coverage), so that the MSIS was aligned with the intended comprehensive coverage definition used in the ACS.⁹

Analysis

We begin by examining aggregated counts from the MSIS and ACS prior to merging. The purpose is to demonstrate the importance of concept alignment by producing counts from the MSIS before and after cleaning procedures that align the MSIS to the ACS concepts of unique persons and comprehensive coverage. Our second goal is to demonstrate that, on an aggregated basis, the ACS, unlike other population surveys, overcounts the MSIS. We argue that this result stems in part from the broad definition of public health insurance used by the ACS and/or false-positive response error.

The second part of the analysis uses the linked file to describe health insurance coverage in the ACS for cases that the MSIS shows have Medicaid or expansion CHIP on the day of interview. We examine two insurance variables. The first indicates if the case was not coded to Medicaid Plus, regardless of other health insurance types reported. It provides the basis for our estimate of the false-negative rate. The second measure includes levels for Medicaid

Plus, no Medicaid Plus but some other type, and uninsured. Using the second variable, we examine the extent that linked cases report other or no insurance.

Using the 2009 ACS/MSIS linked file, we examine disagreements between the ACS and the MSIS for cases that explicitly report Medicaid Plus (i.e., where Medicaid Plus is neither edited nor imputed). We then describe the false-negative rate after sequentially including imputed and edited values.

Next, we examine insurance reports among linked cases by sociodemographics. For that analysis, we use all values (reported or imputed) after all editing procedures have been applied, so that our results reflect data that analysts actually encounter. We then examine false-negative rates across the states, with and without regression adjustments for covariates that vary by state and have an independent association with reporting behavior. Finally, we describe top-level results from the 2008 ACS, which we linked to the MSIS using the same methods described above. We include the 2008 ACS so that we can examine an assumption in the Medicaid evaluation literature: that year-to-year variation in the false-negative rate is constant across the states and does not systematically bias estimates of the impact of Medicaid expansions on health insurance coverage.

Most estimates (with some exceptions) use the poststratified weight. Standard errors are computed using successive difference replication (Fay and Train 1995). Because we have a large sample size, nearly all of our results are statistically significant and we focus our description of results on magnitudes rather than significance tests.

RESULTS

Raw Counts

Prior to aligning the MSIS to the ACS, average daily enrollment in Medicaid, expansion CHIP, and stand-alone CHIP (when reported) recorded by the MSIS totals 50.5 million (Table 1), which nearly matches the ACS Medicaid Plus estimate of 50.6 million (based on the full ACS universe and the original sample weight). The MSIS count includes duplicated and deceased records (2.9 million) and partial benefit plans that are not part of the intended ACS conceptual definition (5.0 million). After removing such enrollees and any stand-alone CHIP cases that are sporadically counted by MSIS, average daily enrollment in the MSIS is 42.6 million. The ACS overcounts this figure by 15.8 percent, but the ACS conceptual definition is broader than the MSIS universe and a large fraction of the discrepancy is likely from stand-alone CHIP.

Table 1: Comparison of Counts from the Medicaid Statistical Information System (MSIS) and the American Community Survey (ACS)

	<i>Count (millions)</i>	<i>Percentage Difference between ACS and MSIS</i>
MSIS count		
All records in MSIS	50.5	0.20
Minus duplicate enrollees and deceased individuals	47.6	5.93
Minus partial benefit enrollees and stand-alone CHIP	42.6	15.81
Plus stand-alone CHIP, per SEDS	46.3	8.50
Original ACS count of Medicaid Plus	50.6	
Linked ACS count		
Weighted number of linked cases	40.0	
Weighted number of linked cases coded to Medicaid Plus	31.4	

Notes. “All records in MSIS” count represents average daily enrollment in 2009 and includes full or partial Medicaid, expansion CHIP, and stand-alone CHIP (where available). “Plus stand-alone CHIP, per SEDS” adds the 3.7 million people enrolled in a stand-alone CHIP on June 30, 2009, as reported in the Statistical Enrollment Data System (SEDS). “Original ACS count” represents Medicaid or other means-tested coverage weighted by the original weight and including all ACS cases. “Linked ACS” counts include only ACS cases with a valid PIK that are linked to MSIS and are shown to be enrolled in full-benefit Medicaid or expansion CHIP on the day of interview. Linked counts use the poststratified weight.

Sources. Medicaid Statistical Information System (MSIS), 2009 American Community Survey (ACS).

The next row adds to the MSIS count 3.7 million people reported to be enrolled in stand-alone CHIP by the Statistical Enrollment Data System (SEDS), the official source of CHIP enrollment (CMS 2012).¹⁰ The ACS overcounts this figure by 8.5 percent. Table 1 illustrates that unlike other population surveys, the ACS appears to overcount rather than undercount Medicaid and CHIP enrollment. The overcount is composed of false-positive survey errors and correct reports for plans not captured in our administrative data. The change in the percentage difference between the ACS and the MSIS, described in the second column, illustrates that there are conceptual alignment issues that render comparisons of raw counts a poor proxy of the survey error in the ACS.

The bottom half of Table 1 displays counts of Medicaid and CHIP enrollment after we merge the ACS and the MSIS. It shows that using the poststratified weight and the full linked ACS file (housing units and group quarters), there are 40 million people that are identified by MSIS as having full-benefit Medicaid or expansion CHIP. This number differs from the count

of 42.6 million found in the MSIS because during the merge we remove MSIS cases that lack a PIK. Despite the apparent overcount in the top half of the table, using linked cases actually coded as Medicaid Plus in the ACS, the estimate is 31.4 million, suggesting that 8.6 million people with known Medicaid or expansion CHIP are not counted by the survey.

Health Insurance Coverage among Linked Cases by Source of Medicaid Plus Value

In Table 2 we report the insurance status of cases that are linked to the MSIS that are in the civilian noninstitutional population. The last two rows show the overall false-negative rate of 21.6 percent or 8.3 million (the change in universe between Tables 1 and 2 causes a change in the false-negative estimate). The top panel describes insurance status prior to the application of logical edits. Among explicitly reported Medicaid Plus values, 22.8 percent do not report Medicaid Plus. Just over 13 percent do not report Medicaid Plus but do report another coverage type, and 9.6 percent are uninsured. After adding imputed cases to explicit reports, the false-negative rate is 24.1 percent. Roughly 14 percent of explicit or imputed cases are coded to another coverage type and 9.9 percent are uninsured. The increase in the false-negative rate between explicit and explicit or imputed cases implies that imputed cases have a higher false-negative rate than reported cases. A higher false-negative rate among imputed cases should not be interpreted as a flaw in the imputation routine. The purpose of imputation is not to correctly code each case but to replicate the distribution of reported values, conditional on the variables included in the imputation routine. While our purpose is not to evaluate the imputation, Table 2 demonstrates that the overall false-negative rate is driven both by explicit reporting and data processing.

The bottom panel of Table 2 describes results after the application of the logical edits. The results suggest that for all values, the edit routine reduces the false-negative rate by 2.5 percentage points. It reduces the uninsured rate among linked cases by 0.8 percentage points. The results imply that the edit routine is moderately successful in reducing false-negatives. However, we are unable to ascertain how many cases the edit routine assigns to Medicaid Plus that do not actually have Medicaid Plus.

Health Insurance Coverage among Linked Cases by Demographic Characteristics

Table 3 presents ACS health insurance rates among known Medicaid and expansion CHIP enrollees across subgroups. As in other surveys, the false-

Table 2: Health Insurance Coverage Estimates from the 2009 ACS, Linked Cases in the Civilian Noninstitutional Universe

	<i>No Medicaid Plus Coverage % (SE)</i>	<i>Any Medicaid Plus Coverage % (SE)</i>	<i>Any Other Coverage % (SE)</i>	<i>Uninsured % (SE)</i>
Prior to logical editing				
Explicit reports only	22.8 (0.13)	77.2 (0.13)	13.2 (0.09)	9.6 (0.09)
Explicit or imputed	24.1 (0.13)	75.9 (0.13)	14.2 (0.09)	9.9 (0.09)
After logical editing				
Explicit reports only	20.3 (0.13)	79.7 (0.13)	11.5 (0.09)	8.8 (0.09)
Explicit or imputed	21.6 (0.12)	78.4 (0.12)	12.5 (0.09)	9.1 (0.08)
Explicit or imputed (pop count)	8.3 m	30.3 m	4.8 m	3.5 m

Notes. Only full-benefit Medicaid or expansion CHIP enrollment from MSIS is counted. The sample size of linked cases is 410,275. Any other coverage includes those that are not coded to Medicaid Plus, but are coded to some other type including ESI, Direct, Medicare, VA, or Tri-Care. SE refers to standard errors from successive difference replication. Explicit Reports versus Imputed values refer only to the values of the Medicaid Plus response item. Ninety-six percent of the sample has reported versus imputed Medicaid Plus.

Source: Reweighted estimates from the 2009 American Community Survey linked to MSIS, Civilian Noninstitutional Population.

negative rate tends to be lower for children, non-Hispanic whites, citizens, and people in lower socioeconomic strata. The right hand panel of Table 3 describes the distribution of coverage types. The uninsured rate among children linked to the MSIS is 7.0 percent versus 14.1 percent for non-elderly adults. For many groups (children, whites, etc.), more linked respondents are coded to another type of coverage versus being coded to uninsured. However, this finding was not universal. Nonelderly adults, American Indian/Alaska Natives, Hispanics, those with less than a high school education, noncitizens, and the unemployed were more likely to be coded as uninsured versus another type of coverage.

False-Negative Rates by State

Figures 1 and 2 map the false-negative rate across the states for children (Figure 1) and adults (Figure 2). We hypothesized that variation between states could be explained by differences in (1) demographic characteristics; (2) program characteristics; (3) the quality of the MSIS; and/or (4) survey operations such as the quality of field representatives. We are unable to assess the role of the last two factors, but below we present evidence on hypotheses 1 and 2.

Table 3: Health Insurance Coverage by Demographics in 2009 ACS, Linked Cases in the Civilian Noninstitutional Universe

	Population Count	No Medicaid Plus		Any Medicaid Plus Coverage % (SE)	Any Other Coverage % (SE)	Uninsured	
		% (SE)	p-value			% (SE)	p-value
Total	38,590,345	21.6 (0.12)		78.4 (0.12)	12.5 (0.09)	9.1 (0.08)	
Gender							
Male	16,902,709	21.3 (0.14)	<.001	78.7 (0.14)	12.4 (0.12)	8.9 (0.11)	.002
Female	21,687,636	21.8 (0.15)		78.2 (0.15)	12.6 (0.11)	9.2 (0.10)	
Age							
0-18	22,038,813	17.7 (0.15)	<.001	82.3 (0.15)	10.8 (0.12)	7.0 (0.10)	<.001
19-64	13,619,349	26.6 (0.18)		73.4 (0.18)	12.5 (0.13)	14.1 (0.14)	
65+	2,932,184	27.0 (0.26)		73.0 (0.26)	25.6 (0.24)	1.4 (0.12)	
Race/ethnicity							
White, non-Hispanic	16,198,649	19.3 (0.15)	<.001	80.7 (0.15)	12.5 (0.13)	6.9 (0.10)	<.001
Black, non-Hispanic	9,358,265	22.0 (0.22)		78.0 (0.22)	13.9 (0.19)	8.1 (0.15)	
AIAN, non-Hispanic	626,674	24.4 (0.88)		75.6 (0.88)	9.5 (0.45)	14.9 (0.71)	
Asian/NHOPI, non-Hispanic	1,630,923	28.1 (0.50)		71.9 (0.50)	20.0 (0.49)	8.1 (0.36)	
Other/Mult, non-Hispanic	115,089	23.1 (1.88)		76.9 (1.88)	16.4 (1.47)	6.7 (1.10)	
Hispanic, any race	10,660,746	23.4 (0.22)		76.6 (0.22)	10.3 (0.15)	13.1 (0.19)	
Education of household head							
LT high school	13,363,113	20.0 (0.19)	<.001	80.0 (0.19)	9.7 (0.14)	10.3 (0.15)	<.001
HS or GED	11,479,509	21.0 (0.22)		79.0 (0.22)	11.9 (0.16)	9.1 (0.14)	
More than HS	13,747,724	23.6 (0.20)		76.4 (0.20)	15.7 (0.15)	7.9 (0.13)	

continued

Table 3: *Continued*

	Population Count	No Medicaid Plus		p-value	Any Medicaid Plus Coverage % (SE)	Any Other Coverage % (SE)	Uninsured	
		% (SE)	p-value				% (SE)	p-value
Employment of household head								
Not in labor force	32,206,973	19.4 (0.14)	<.001	80.6 (0.14)	11.6 (0.10)	7.8 (0.08)	<.001	
Unemployed	1,133,756	26.3 (0.63)		73.7 (0.63)	10.1 (0.40)	16.3 (0.46)		
Employed	5,249,616	33.7 (0.32)		66.3 (0.32)	18.5 (0.24)	15.1 (0.25)		
Citizenship								
Non-citizen	1,871,750	38.4 (0.51)	<.001	61.6 (0.51)	14.2 (0.36)	24.2 (0.50)	<.001	
Citizen	36,718,595	20.7 (0.13)		79.3 (0.13)	12.4 (0.09)	8.3 (0.08)		
Poverty level								
0-138	24,063,298	15.8 (0.13)	<.001	84.2 (0.13)	7.4 (0.09)	8.4 (0.10)	<.001	
139-249	8,518,181	28.0 (0.26)		72.0 (0.26)	17.3 (0.21)	10.8 (0.19)		
250-399	3,541,868	35.7 (0.38)		64.3 (0.38)	25.5 (0.33)	10.2 (0.23)		
400+	1,932,740	39.1 (0.55)		60.9 (0.55)	30.2 (0.48)	8.9 (0.32)		

Notes: In this table, we include explicit and imputed values after the application of logical edits. The weighted population count in this table does not agree with the linked count in Table 1 because this table is restricted to the civilian noninstitutional population. Only full-benefit Medicaid or expansion CHIP enrollment from MSIS is counted. Any other coverage includes those that are not coded to Medicaid Plus, but are coded to some other type including ESI, Direct, Medicare, VA, or Tri-Care. SE refers to standard errors from successive difference replication. *p*-values are from Wald tests.

Source: Reweighted estimates from the 2009 American Community Survey linked to MSIS, Civilian Noninstitutional Population (*n* = 410,275).

Panel A of Figure 1 shows unadjusted results for children (see values in Table S4a). The unadjusted false-negative rate ranged from 10.5 percent in Maine to 25.5 percent in Nevada. Panel B shows adjusted rates, controlling for gender, race, poverty, citizenship, household education, and employment. If demographics explain the variation shown in Panel A, Panel B would show a single color. However, the model only moderately reduces variation across the states. Table S4b shows that the standard deviation of the unadjusted rates (3.3) is minimally reduced after adjustment (3.2), and the range remained unchanged (15.0 percentage points). The unadjusted false-negative rates for adults, shown in Panel A of Figure 2, ranged from 15.1 percent in Vermont to 34.2 percent in Arizona. Adjustment (Panel B) had the same minimal effect on variation between the states as observed for children.

To examine hypothesis 2, we present results from a logistic model of false-negative reports regressed on demographics and state program characteristics in Table S6. We controlled for program name complexity and relative income eligibility thresholds for children and parents (low, medium, or high). We also controlled for whether childless adults were covered, if the state had only a stand-alone CHIP and presumptive eligibility policies. None of the coefficients suggest that program characteristics had a meaningful impact on false-negative error.

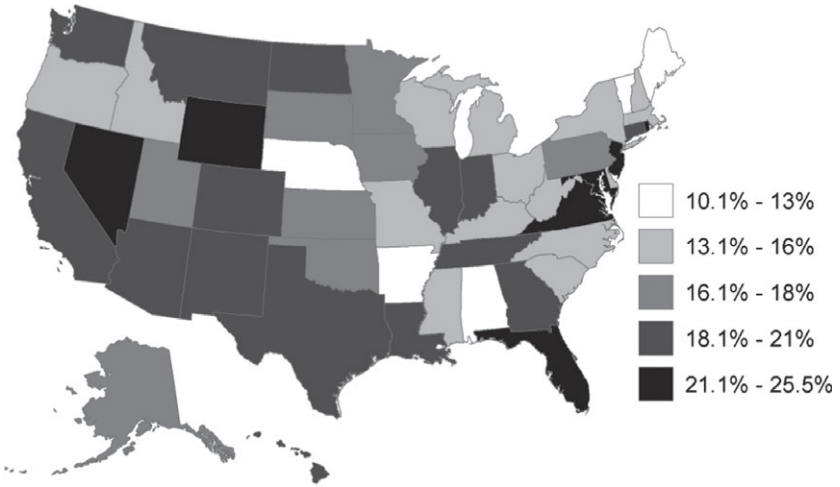
Medicaid Plus Survey Error and Uninsurance

Table 3 shows that 9.1 percent of known Medicaid/expansion CHIP enrollees were coded as uninsured. In data not presented here, we found that 7.7 percent of the uninsured were enrolled in Medicaid or expansion CHIP, per MSIS. This suggests that the 45.4 million people estimated to be uninsured is upwardly biased by Medicaid/expansion CHIP misreporting by 3.5 million people (7.7 percent). The upward bias is likely offset by people coded to Medicaid Plus who are actually uninsured. Furthermore, there is little information in the literature on the effect of other coverage type misreporting on uninsurance. Due to these limitations, our estimate of the bias to uninsurance (3.5 million) should be interpreted with extreme caution.

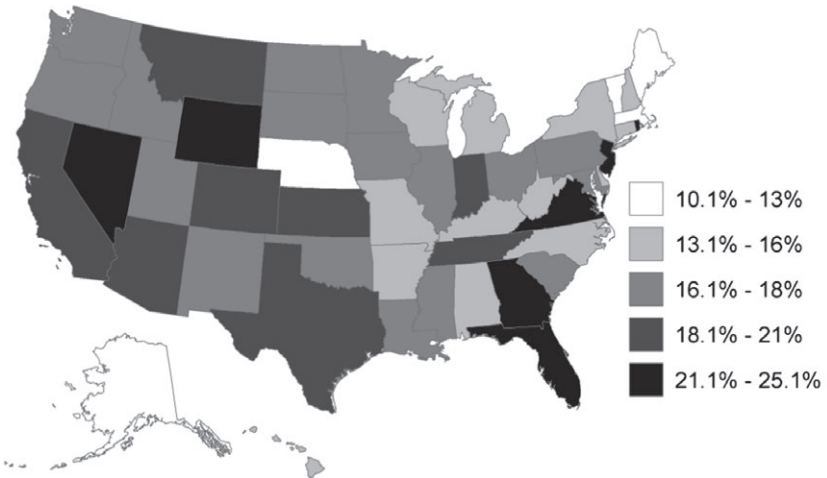
Medicaid Survey Error by State and Year

An assumption in difference-in-differences studies that measure the impact of Medicaid/CHIP expansions on health insurance coverage is

A



B



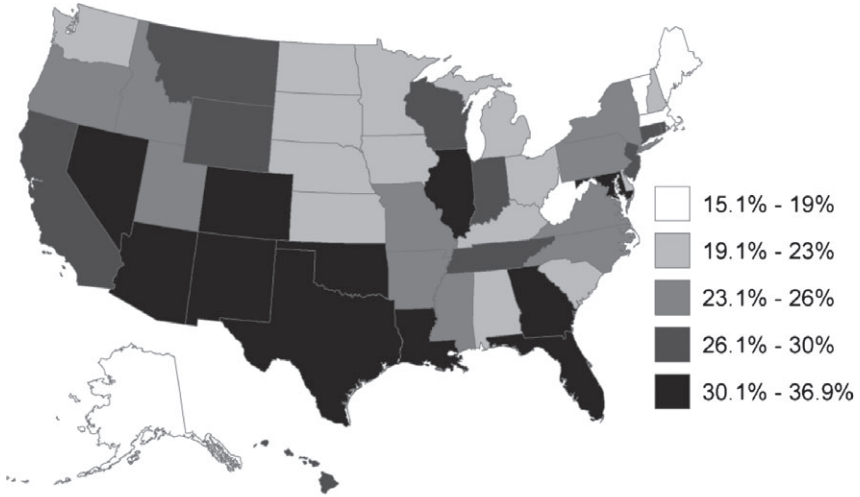
that year-over-year changes in measurement error are independent of state. If a change in measurement error is correlated with the expansion, the results will be systematically biased. In the appendix we present an initial evaluation of this assumption using linked 2008 and 2009 ACS/MSIS data.¹¹ At the national level, the false-negative rate in the 2008 ACS was 24.1 percent, statistically different than the 2009 rate (21.6 percent).

To examine state-by-year-variation, we appended the files and ran nested logistic-regression models on linked cases. The first block of variables included demographics and year, the second block included state, and the third block included the interaction of state and year. We examined two dependent variables: no Medicaid Plus and uninsurance. Each block was statistically significant ($p < .001$) in the Medicaid Plus regression (Table S7). In the uninsured regression, all but the block of interaction effects was statistically significant (Table S8). The results from the Medicaid plus regression do not specifically indicate if the change in the false-negative rate is correlated with a change in eligibility, but the statistical significance on the block of interaction effects, in combination with the correlation between false-negative error and poverty (Table 3), suggests that the assumption is suspect. This finding did not carry over to uninsurance. This suggests a differential change in how respondents classified their type of coverage, but not their insured status. The lack of a significant finding on the block of interaction effects provides support for the assumption that the impact of false-negatives on insured/uninsured estimates does not vary at the state-by-year level. However, because we only have 2 years of data, this finding should be interpreted with caution.

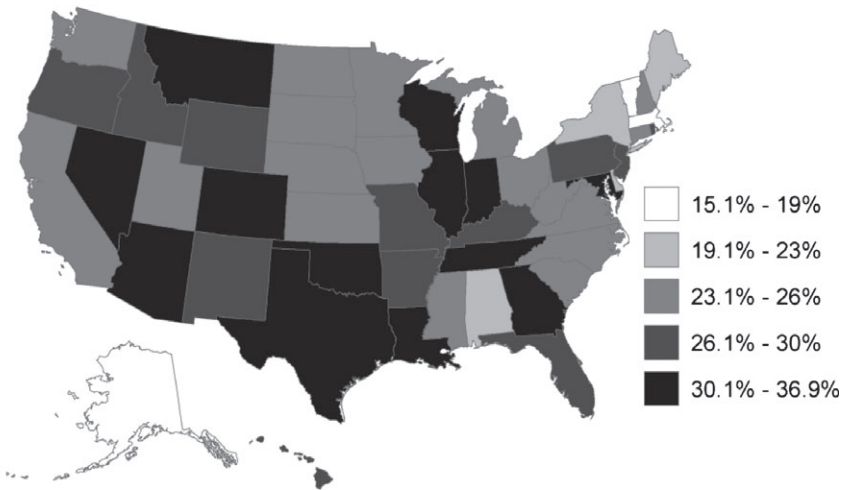
Figure 1: Unadjusted and Adjusted False-Negative Rates for Children. (A) Unadjusted Rate for Children Age 0–18. (B) Adjusted Rate for Children Age 0–18

Source: Reweighted estimates from the 2009 American Community Survey linked to MSIS, civilian noninstitutional population. The regression adjusted rates are based on a logistic regression model and are calculated using average marginal effects. They control for race, gender, poverty status, citizenship, household education, and household employment. Tables of values and regression coefficients are included in the appendix.

A



B



DISCUSSION

This article examined measurement error in public health insurance in the ACS. Our study had limitations. The quality of the MSIS varies across the states and the lack of stand-alone CHIP data in the MSIS could be particularly problematic for making comparisons across states. However, if our results are only generalized to Medicaid and expansion CHIP, then comparisons across states would only be biased if Medicaid enrollees (and expansion CHIP enrollees in expansion and combination states) reporting varied by the state's CHIP type-decision. In a regression analysis, we did not find evidence of such a difference, which gives us confidence that the lack of stand-alone CHIP data did not bias our examination of state variation in false-negative error. Beyond CHIP, the MSIS lacks information on other non-Medicaid programs that fall within the scope of the ACS question, and we were unable to confidently estimate false-positive errors. Another potential concern is whether MSIS data quality varies by year, in which case our state-by-year analysis would be biased. Finally, missing PIKs in the MSIS could have caused bias to our estimate of false-negatives if non-PIK'ed cases had different reporting patterns than PIK'ed cases.

Table 1 showed that the overall count in the ACS is larger than the count from the MSIS, even after we edit the MSIS to mimic the universe of the ACS and add enrollment in stand-alone CHIP. This is a unique finding compared to other federal surveys that have been compared to the MSIS. The overcount is caused by the concept misalignment between the ACS question and MSIS enrollment and/or false-positive survey error. Unfortunately, we lacked administrative counts for the full scope of plans covered by the intended ACS definition and are unable to determine if the count in the ACS under or over-estimates its intended target. However, if data users wish to interpret the ACS

Figure 2: Unadjusted and Adjusted False-Negative Rates for Adults. (A) Unadjusted Rate for Adults Age 19 and over. (B) Adjusted Rate for Adults Age 19 and over

Source: Reweighted estimates from the 2009 American Community Survey linked to MSIS, civilian noninstitutional population. The regression adjusted rates are based on a logistic regression model and are calculated using average marginal effects. They control for race, gender, poverty status, citizenship, household education, and household employment. Tables of values and regression coefficients are included in the Appendix.

item as Medicaid or CHIP (not including any state-specific plans), as is often done in take-up studies (e.g., Kenney et al. 2010), then our results suggest that the ACS does a relatively adequate job without the need for additional data processing edits.

Using the ACS-MSIS linked file, we found that 78.4 percent of known Medicaid and expansion CHIP enrollees are coded to Medicaid Plus. Similar to other surveys, we found that the level of false-negatives varied across groups. In general, children and groups with lower socioeconomic standing tended to have lower levels of false-negative error. This could be because such groups are more closely tied to public programs and are more knowledgeable about their enrollment. For many groups, miscoded enrollees were coded into other coverage types rather than as uninsured, but there were important exceptions.

We found that the false-negative rate varied across the states. We were unable to attribute any meaningful portion of the observed variation to specific causes. Variation remained after controlling for demographics, and we did not identify any meaningful program characteristics. However, we lacked important information about the nature of enrollment. It could be that a higher level of underreporting in some states arises not because respondents do not understand the question, but because they were not adequately educated about their benefits during enrollment. Evaluations of the CPS and NHIS suggest that Medicaid enrollees that are coded as uninsured tend to utilize fewer services than their counterparts (University of Minnesota's State Health Access Data Assistance Center SNACC Phase II, SNACC Phase IV). This finding is consistent with the idea that undercounted cases are not aware of their enrollment and could use services in a suboptimal way. We were also unable to estimate the impact of divergent MSIS quality or survey operations on false-negative reports.

The level of false-negative error in the ACS is discouraging, but the ACS compares favorably to other surveys. Comparisons across surveys should be interpreted with caution. We made every attempt to follow the procedures of other record check evaluations, but the methods were not identical. With this caveat in mind, the false-negative rate in the 2005 CPS was 42.9 percent and in the 2002 NHIS was 32.2 percent. The 2009 ACS false-negative rate of 21.6 percent is most similar to the MEPS false-negative rate of 17.5 percent. The relative proximity of the MEPS and ACS could be, in part, due to the designation of CHIP and other public programs as "correct" reporting in both ACS and MEPS, but not CPS or NHIS. Another comparison that avoids confounding how other programs are treated is the rate of uninsurance among known enrol-

lees. In the 2005 CPS, 17.9 percent of known enrollees were coded as uninsured, versus 8.8 percent of enrollees in the 2002 NHIS and 8.3 percent in the 2003 MEPS. Using the uninsured rate also suggests that the 2009 ACS (9.1 percent) is on the lower end of the false-negative continuum.

In the context of other surveys, our results suggest data users are relatively well-served by the ACS. This is an important finding because it suggests that adding complexity to an instrument (i.e., requiring respondents produce a health insurance card) may not necessarily improve on the quality of estimates that can be obtained from a simple health insurance measure. This research was not designed to identify the exact survey design features in the ACS that lead to response error. Future research is needed, preferably using a split-ballot experiment that varies questionnaire features and is based on a sample with known health insurance types. Such a design could estimate the level of ACS survey error in other coverage types and lead to strategies that improve the ACS.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This project was funded by a contract from the U.S. Census Bureau to the State Health Access Data Assistance Center (SHADAC). Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau. We thank Karen Turner at SHADAC for GIS support and Rachelle Hill of the Census Bureau for assistance at the Minnesota Research Data Center. We thank the editors and two anonymous reviewers for helpful comments that improved the manuscript.

Disclosures: None.

Disclaimers: None.

NOTES

1. State-specific plans include programs operated within states that do not receive matching federal funds. These programs include high-risk pools, programs for low-income persons that are not eligible for Medicaid, and county-specific programs.
2. Starting in 2012, the ACS also provided an Internet response option.
3. In Table S3, we present results for housing units, noninstitutional group quarters, and institutional group quarters. These results show that the percent of false-negatives in group quarters is slightly lower than reporting in housing units and that

there is very little variation between institutional and noninstitutional residents. However, residents in housing units are twice as likely to be coded as uninsured, compared to residents in group quarters.

4. The ACS gathers information on participation in the Indian Health Service, but this coverage type is not considered comprehensive and is not used when defining insured.
5. Other surveys distinguish between Medicaid, CHIP, and state-specific programs, but the recommended practice is to combine these variables into a single composite variable (e.g., MACPAC 2012).
6. We estimate that 7.0 percent of cases that are not found on the MSIS report Medicaid Plus coverage (Table S9).
7. In 2009, stand-alone CHIP states included Alabama, Arizona, Colorado, Connecticut, Georgia, Kansas, Mississippi, Montana, Nevada, New York, Oregon, Pennsylvania, Texas, Utah, Vermont, Washington, West Virginia, and Wyoming. CHIP expansion states included Alaska, DC, Hawaii, Maryland, Nebraska, New Mexico, and Ohio. Combination CHIP states included Arkansas, California, Delaware, Florida, Idaho, Illinois, Indiana, Iowa, Kentucky, Louisiana, Maine, Massachusetts, Michigan, Minnesota, Missouri, New Hampshire, New Jersey, North Carolina, North Dakota, Oklahoma, Rhode Island, South Carolina, South Dakota, Tennessee, Virginia, and Wisconsin.
8. In Table S2, we show that our estimate of the false-negative rate is slightly higher using the poststratified versus the original weight.
9. The inclusion of pregnancy coverage aligns our analysis with previous record check studies. Analysis of claims data for those enrolled for pregnancy related benefit indicated patterns of service utilization similar to full benefits Medicaid (University of Minnesota's State Health Access Data Assistance Center SNACC Phase II).
10. Because we did not have direct access to the SEDS microdata, we could not ensure that it was cleaned in the same manner as our MSIS file. To our knowledge, stand-alone CHIP funds cannot be used to finance partial-benefit plans, but in some instances they may be used for wrap-around coverage.
11. For more detailed results from the 2008 linked ACS/MSIS file, see Boudreaux et al. (2013).

REFERENCES

- Abraham, J. M., P. Karaca-Mandic, and M. Boudreaux. 2013. "Sizing Up the Individual Market for Health Insurance: A Comparison of Survey and Administrative Data Sources." *Medical Care Research and Review* 70 (4): 418–33.
- Blewett, L. A., and M. Davern. 2006. "Meeting the Need for State-Level Estimates of Insurance Coverage: Use of State and Federal Survey Data." *Health Services Research* 41 (3): 946–75.
- Boudreaux, M., J. Y. Ziegenfuss, P. Graven, M. Davern, and L. A. Blewett. 2011. "Counting Uninsurance and Means-Tested Coverage in the American Commu-

- nity Survey: A Comparison to the Current Population Survey.” *Health Services Research* 46 (1): 210–31.
- Boudreaux, M., K. T. Call, J. Turner, B. Fried, and B. O’Hara. 2013. “Accuracy of Medicaid Reporting in the ACS: Preliminary Results from Record Linkage.” U.S. Census Bureau Working Paper [accessed on December 1, 2013]. Available at http://www.shadac.org/files/shadac/publications/ACSUndercount_Working_Paper_0.pdf
- Boudreaux, M., K. T. Call, J. Turner, and B. Fried. 2014. *Estimates of Direct Purchase from the ACS and Medicaid Misreporting: Is There a Link?*. SHADAC Brief #38. Minneapolis, MN: State Health Access Data Assistance Center.
- Call, K. T., M. E. Davern, J. A. Klerman, and V. Lynch. 2012. “Comparing Errors in Medicaid Reporting across Surveys: Evidence to Date.” *Health Services Research* 48 (2 Pt 1): 652–64.
- Centers for Medicare and Medicaid Services [CMS]. 2012. *FY 2009 Third Quarter – Program Enrollment Last Day of Quarter by State – Total CHIP* [accessed on December 1, 2013]. Available at <http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Childrens-Health-Insurance-Program-CHIP/CHIP-Reports-and-Evaluations.html>
- Centers for Medicare and Medicaid Services [CMS]. 2013a. *MSIS State Data Characteristics/Anomalies Report* [accessed on December 1, 2013]. Available at http://www.cms.hhs.gov/medicaidDataSourcesGenInfo/02_MSISData.asp
- Centers for Medicare and Medicaid Services [CMS]. 2013b. *Medicaid Data Sources—General Information* [accessed on December 1, 2013]. Available at http://www.cms.hhs.gov/medicaidDataSourcesGenInfo/02_MSISData.asp
- Czajka, J. 2012. “Movement of Children between Medicaid and CHIP, 2005-2007.” Mathematica Policy Research Brief 4, March 2012 [accessed on December 1, 2013]. Available at http://www.mathematica-mpr.com/publications/PDFs/health/medicaid_chip_05-07.pdf
- Czajka, J. L., and S. Verghese. 2013. “Social Security Numbers in Medicaid Records: Reporting and Validity, 2009.” Mathematica Policy Research, Inc. [accessed on December 1, 2013]. Available at <http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/Downloads/FinalSSNreport.pdf>
- Davern, M., B. Quinn, G. Kenny, and L. A. Blewett. 2009a. “The American Community Survey and Health Insurance Coverage Estimates: Possibilities and Challenges for Health Policy Researchers.” *Health Services Research* 44 (2): 593–605.
- Davern, M., J. A. Klerman, D. K. Baugh, K. T. Call, and G. D. Greenberg. 2009b. “An Examination of the Medicaid Undercount in the Current Population Survey: Preliminary Evidence from Record Linkage.” *Health Services Research* 44 (3): 965–87.
- Fay, R. E., and G. F. Train. 1995. *Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties*. Proceedings of the American Statistical Association Conference, Orlando, FL.

- Kenney, G., V. Lynch, A. Cook, and S. Phong. 2010. "Who and Where are the Children Yet to Enroll in Medicaid and the Children's Health Insurance Program?" *Health Affairs* 29 (10): 1920–9.
- Klerman, J. A., M. R. Plotzke, and M. Davern. 2012. "CHIP Reporting in the CPS." *Medicare and Medicaid Research Review* 2 (3): E1–14.
- Klerman, J. A., M. Davern, K. T. Call, V. Lynch, and J. D. Ringle. 2009. "Understanding the Current Population Survey's Insurance Estimates and The Medicaid 'Undercount'." *Health Affairs* 28 (6): w991–1001.
- Lynch, V., M. Boudreaux, and M. Davern. 2009. "Applying and Evaluating Logical Coverage Edits to Health Insurance Coverage in the American Community Survey." U.S. Census Bureau Working Paper [accessed on December 1, 2013]. Available at https://www.census.gov/hhes/www/hlthins/publications/coverage_edits_final.pdf
- Mach, A., and B. O'Hara. 2011. "Do People Really Have Multiple Health Insurance Plans? Estimates of Nongroup Health Insurance in the American Community Survey." SEHSD Working Paper Number 2011-28 [accessed December 1, 2013] Available at https://www.census.gov/hhes/www/hlthins/publications/Multiple_Coverage.pdf
- Medicaid and CHIP Payment and Access Commission [MACPAC]. 2012. "2011 Medicaid and CHIP Program Statistics: MACStats" [accessed on December 1, 2013]. Available at http://www.macpac.gov/macstats/MACStats_March_2011.pdf?attredirects=0&d=1
- Mulrow, E., A. Mushtaq, S. Pramanik, and A. Fontes. 2011. *Assessment of the U.S. Census Bureau's Person Identification Validation System*. Bethesda, MD: NORC at the University of Chicago.
- Pascale, J. 2008. "Measurement Error in Health Insurance Reporting." *Inquiry* 45 (4): 422–37.
- Turner, J., M. Boudreaux, and V. Lynch. 2009. "A Preliminary Evaluation of Health Insurance Coverage in the 2008 American Community Survey." U.S. Census Bureau Working Paper [accessed on December 1, 2013] Available at http://www.census.gov/hhes/www/hlthins/data/acs/2008/2008ACS_healthins.pdf
- University of Minnesota's State Health Access Data Assistance Center, the Centers for Medicare and Medicaid Services, the Department of Health and Human Services Assistant Secretary for Planning and Evaluation, The National Center for Health Statistics, and the U.S. Census Bureau (SNACC). 2001–2005. "Phase II-Phase VI Research Results" [accessed on December 1, 2013]. Available at <http://www.census.gov/did/www/snacc/>
- U.S. Census Bureau. 2014. "Response Rates – Data" [accessed on November 24, 2014]. Available at http://www.census.gov/acs/www/methodology/response_rates_data/
- Wagner, D., and M. Layne. 2012. *The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software* [accessed on January 15, 2014]. Available at http://www.census.gov/srd/carra/CARRA_PVS_Record_Linkage.pdf

SUPPORTING INFORMATION

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

Appendix SA1: Author Matrix.

Appendix SA2: Technical Appendix.