Medicaid Enrollment: The Relationships between Survey Design, Enrollee Characteristics, and False-Negative Reporting

Victoria Lynch
US Census Bureau, 4600 Silver Hill Road, Suitland, MD 20233
victoria.a.lynch@census.gov

Abstract
Record check studies show that survey estimates of enrollment in government-assistance programs tend to be lower than those compiled from records used for program administration, and this undercount is especially apparent for Medicaid. Studies specific to Medicaid point to false-negative reporting about enrollees in surveys as the main explanation, however their results differ with respect to findings about the level of this response error. It is unclear how much study differences owe to genuine discrepancies in how different surveys measure Medicaid versus being artifacts of different methods for measuring the undercount. This study helps to clarify this question by comparing the results of using one set of variables, derived from the same administrative database, to separately model Medicaid misreport in two different surveys: the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) and in the National Health Interview Survey (NHIS), both fielded in 2001.

Results suggest that survey design has an important effect on Medicaid reporting, and most notably that differences in reference period are enough to explain differences in the probability of false-negative reporting in CPS and NHIS, controlling for sample differences. Results corroborate findings that the probability of misreport depends on particular characteristics of the enrollee, and also add evidence suggesting that many enrollee-related predictors of misreport may be quite robust to some differences in survey design. To learn more about the relationships between specific features of survey design and false negative reporting, these results imply that it would be fruitful to look closer at the effect of having private insurance (at the same time as Medicaid) and of being enrolled in other assistance programs.

Key Words: Medicaid undercount, administrative records, CPS, NHIS, misreport, false-negative reporting

This report is released to inform interested parties of ongoing research and to encourage discussion. Any views expressed are those of the author and not necessarily those of the U.S. Census Bureau or any other institution.

1. Background

This paper reports about a study to evaluate discrepancies between survey estimates of enrollment in Medicaid and the number of enrollees reported in state and national administrative data. Studies using different surveys show that discrepancies range from about 10% to 38% (Blumberg & Cynamon, 1999; Call, Davidson, Sommers, Feldman & Rockwood, 2002; Card, Hildreth, & Shore-Sheppard, 2001; Czajka & Lewis, 1999; Eberly, Pohl, & Davis, 2005; Klerman, Roth, & Ringel, 2005; Lewis, Ellwood, & Czajka, 1998; SNACC, 2008). It is unclear how much this range owes to genuine discrepancies in how different surveys measure Medicaid as opposed to being an artifact owing to the use of different methods for measuring the undercount. By contrast, studies are quite consistent in their finding that surveys undercount Medicaid largely because respondents misclassify enrollment status (typically reporting some non-Medicaid form of coverage) (Call, Davidson, Davern, & Nyman, 2008; Davern, Call, Beebe, Bland, Ziegenfuss, & Blewet, 2008; Eberly et al, 2005; Klerman et al, 2005; SNACC, 2008). This response error is largely problematic because it appears to be non-random (e.g., respondents are less likely to misclassify lower income enrollees than higher income enrollees), meaning that the cases identified as enrollees may not be a random subset of the sample truly enrolled. This may result in bias in the distribution of program funds (formulas use survey data) as well as misleading results in analyses that rely on survey responses about Medicaid.

Despite the scope of possible problems from the Medicaid undercount, there is reason for optimism if surveys can be modified to improve the reporting of Medicaid enrollment. Evidence suggests it is technically possible. For example, there is evidence that the choice of reference period relates to the probability of correct reporting (SNACC, 2008). There is also evidence that specific characteristics of a person's experience in Medicaid relate to correct reporting, so it is possible that this knowledge can be used to develop survey questions that make enrollment more apparent to respondents. However, there is little evidence about the practical consequences of such modifications. For example, little is known about the relationships between various features of survey design and reporting Medicaid. Without that knowledge it is difficult to assess the relative costs and benefits of possible modifications, both in terms of measuring Medicaid enrollment and in the context of a given survey's broader purposes.

2. Research Question and Study Methods

The purpose of this research was to study how survey design and enrollee characteristics affect false-negative reporting of Medicaid. Methods included comparing patterns of false-negative reports across survey conditions, and attributing observed differences to the possible effects of the design features distinguishing the survey conditions.

2.1 The Research Files and Record Check

The study was conducted using a file of anonymized person-level[1] records from NHIS and CPS linked[2] to matching records (i.e., for the same individual) found in the Medicaid Statistical Information System (MSIS).[3] False-negative cases were identified as any case found in MSIS with a record for comprehensive Medicaid coverage during the survey's reference period but no survey report of Medicaid enrollment (reported by the enrollee or by proxy). The reference period for CPS is the previous calendar year (2000), a period that ended about three months prior to the date of interview. In contrast, the reference period for NHIS is the date of the interview in 2001.

[1] All data manipulation and analysis was done in accordance with the Census Bureau's standards for maintaining data security and protecting individual privacy.
[2] See www.census.gov/did/www/snacc for more information about the method used for linking.
[3] MSIS is a database containing enrollment records used by states to administer their Medicaid programs.

2.2 The Logistic Regression Model and Comparison of Regression Coefficients

The analysis was conducted using logistic regression[4] to identify enrollee characteristics predictive of misreport and also to evaluate the effect of survey design on misreport. The model (see Figure 1) used a single set of predictor variables derived from demographic, economic and other enrollee information (mostly from MSIS regardless of survey condition). Design features were evaluated by testing for equality[5] of regression coefficients (for the same predictor variable) estimated using subsets from different reporting conditions. Tests comparing results from conditions 2 and 3 were used to evaluate recall period (because the CPS subsets differed with respect to whether or not the respondent had to recall past enrollment to correctly report enrollment). Tests comparing conditions 1 and 2 were used to evaluate CPS-NHIS design differences other than recall period (because although the subsets were in surveys with different reference periods, there should be no opportunity for the CPS retrospective reference period to affect reporting since the CPS subset was enrolled during the reference period and the date of interview).

Figure 1. The Model

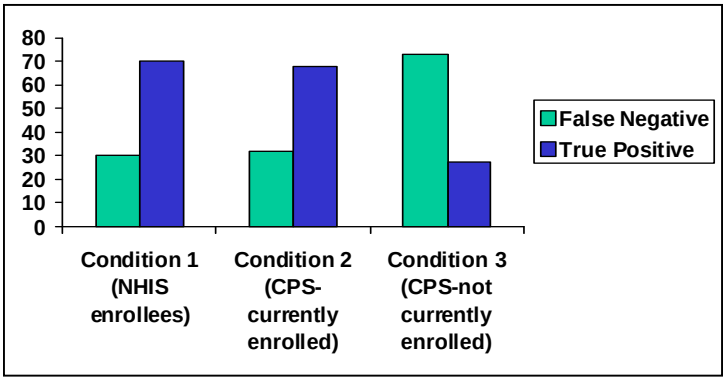
Pr(false-negative report) = f(enrollee characteristics)
Condition 1: NHIS | current enrollee Condition 2: CPS | current enrollee Condition 3: CPS | not current enrollee

3. Results

3.1 Rates of False-Negative Reporting

Overall false-negative reporting about sample enrollees appears to have been less prevalent in NHIS than in CPS. Thirty percent of NHIS respondents did not report Medicaid enrollment but were present in MSIS, and forty percent of CPS respondents had a false-negative report. [6] However, Figure 2 shows that the rates were similar in conditions 1 and 2 (30.1% and 32.1% respectively), when the samples of respondents for analysis were restricted to those reporting about enrollees who had Medicaid coverage at the time of the CPS or NHIS interview. The level of misreport was 72.81% in condition 3, when CPS respondents were reporting about enrollees who were not covered at the time of the interview.

Figure 2. Accuracy of Reporting about Medicaid Enrollees, NHIS and CPS responses collected in 2001



3.2 Logistic Regression Results from Modeling Misreport of Enrollees in Different Survey Conditions

Regression results corroborate previous findings that the following enrollee characteristics are good predictors of misreport: demographics, type of eligibility for Medicaid, and experience with Medicaid services (Klerman et al., 2005; SNACC, 2008). Table 1 shows the enrollee variables found to have a

[4] SAS's survey logistic procedure and sampling weights were used to account for sample design.
[5] The formula for the test statistic is as follows: (b1-b2)/SQRT(SE(b1)**2 + SE(b2)**2).
[6] Reporting rates were calculated without weights because the focus of this study is the effect of the surveys on the sample respondents—not to make inference about the proportion of the total Medicaid population (within the sampling frames) that would be false-negative reports in NHIS or CPS.

statistically significant relationship with false-negative report. Results showed no discernible patterns[7] for the state variables (e.g., only Washington had a significant effect in the same direction across conditions 1 and 3) and given page limits, those results are reported on www.census.gov/did/www/snacc.

Table 1. Results from the Logistic Model of False-Negative Reporting, NHIS and CPS Responses Collected in 2001

Variable	NHIS Current Enrollees	CPS Current Enrollees	CPS Not Current Enrollees
Intercept	1.3708**	1.3691**	2.2698**
Age: 0-5	-0.2765**	-0.3355**	-0.3522
6 – 14	-0.1315	-0.1839	-0.2522
15 – 17	-0.1697	-0.1565	-0.0863
18 – 44	-0.0671	0.1382	0.2468
45 – 64	0.00225	-0.1015	0.3854
65+	0.6466**	0.6392**	0.0585
Hispanic or Minority	0.1721	0.3292**	0.4787*
Male	0.0171	-0.0153	-0.1363
Relationship to Reference Person: Self	-0.00224	-0.1209	-0.1869
Parent	-0.1232	-0.00819	0.3285
Spouse	0.2170	0.0802	-0.1989
Own Child	0.0504	-0.1590	-0.0701
Other	-0.1420	0.2078*	0.1274
Income-to-Poverty Ratio: < 50%	-0.5745**	-0.5081**	-0.7427**
50-74%	-0.3482	-0.5003	-0.2596
75-99%	-0.4175	-0.1193	-0.1573
100-124%	0.1382	-0.0813	0.0652
125-149%	0.0775	0.1353	0.1345
150-174%	0.1818	0.1329	0.2836
175-199%	0.3465	0.2792	0.1976
> 199%	0.5819**	0.6615**	0.4788**
Medicaid Supplements Medicare	-0.4292*	-0.3904**	-0.8540*
Medicaid with Private Insurance	0.6511**	0.1567	-1.0633
Medicaid Due to TANF coverage	-0.3358*	-0.3962**	0.0924
SSI	-0.3908**	-0.7678**	-0.3487
Days with Coverage in 2000	-0.00151**	-0.00283**	-0.00362**
Payment for Service in 2000	-0.5017**	-0.5201**	-0.5333**
Payment for Prescription: Last 30 days	-0.6086**	-0.3378**	-1.4114**
31-60 days ago	-0.5261**	-0.2873**	-0.6535
61-90 days ago	-0.1214	-0.2108	-0.6929
Payment for Non-Prescription Service: Last 30 Days	-0.9759**	-0.4907**	-0.7339**
31-60 days ago	-0.5426**	-0.5262**	-0.2404
61-90 days ago	-0.5423*	-0.4028*	-0.3072
State	See www.census.gov/did/www/snacc		

** indicates a p-value less than .01
* indicates a p-value less than .05

[7] Finding no discernible pattern for state adds credibility to the use of MSIS data for checking survey response because the presence of any endemic errors in a given state's administrative enrollment data would be expected to yield positive relationships with false-negative reporting in each survey condition.

3.3 Tests of Equality of Regression Coefficients Estimated from Different Survey Conditions
Table 2 shows statistically significant results of comparison tests of model coefficients along with one notable non-significant result. The non-significance for the test of difference in intercepts for conditions 1 versus 2 is notable because it indicates that the probability of misreporting an enrollee with current coverage was about the same for NHIS and CPS, controlling for the enrollee characteristics in the model. The results also show that relatively few of the enrollee characteristics had a different relationship with false-negative reporting in the different conditions. For example, Table 2 shows that only seven of the coefficients listed in Table 1 (which excludes state) differ for NHIS and CPS current, and three of those should differ if respondents correctly attend to the respective reference period (e.g., days with coverage in 2000 is outside the NHIS reference period).

Table 2. Notable P-Values from Tests of Equality of Beta Coefficients

Coefficient Tested		Conditions 1 Versus 2 (Current Enrollees in NHIS and CPS)	Conditions 2 Versus 3 (Current and Non-Current Enrollees in CPS)
Intercept		ns	0.0017
Age 18 – 44		0.0424	ns
Medicaid with Private Insurance		0.0257	ns
"Other" Relationship to Reference Person		0.0190	ns
Medicaid Due to TANF Coverage		ns	0.0248
SSI		0.0114	ns
Payment for Prescription: Last 30 days		0.0380	0.0227
Payment Non-Prescription Service: Last 30 Days		0.0045	ns
Days with Coverage in 2000		0.0064	ns
State:	Arizona	0.0194	ns
	Colorado	0.0437	ns
	Connecticut	ns	0.0168
	Illinois	0.0272	ns
	Indiana	0.0003	ns
	Massachusetts	0.0048	ns
	Minnesota	0.0016	ns
	New Hampshire	0.0465	ns
	New York	0.0278	ns
	North Carolina	0.0545	0.0398
	Ohio	0.0021	ns
	Oklahoma	0.0003	ns
	Oregon	0.0010	ns
	Rhode Island	<.0001	0.0340
	South Dakota	<.0001	ns
	Tennessee	<.0001	ns

Table 2. Notable P-Values from Tests of Equality of Beta Coefficients		
Coefficient Tested	**Conditions 1 Versus 2** (Current Enrollees in NHIS and CPS)	**Conditions 2 Versus 3** (Current and Non-Current Enrollees in CPS)
Texas	0.0018	ns
Virginia	ns	0.0135
Washington	0.0453	ns
Wisconsin	0.0296	ns

"ns" indicates a non-significant test result

4. Discussion

4.1 Main Findings

1) Figure 2 shows that false-negative reporting was common across the survey conditions.
2) Table 1 shows that most enrollee characteristics were predictors in multiple conditions.
3) Table 2 shows that the relationships between enrollee characteristic and report tended to be similar across the different survey conditions (i.e., few significant differences in betas).
4) Table 2 shows that recall period can account for the different probabilities of misreport in CPS and NHIS (i.e., no significant difference in the intercepts for current enrollees).
5) Table 2 shows that reference period is not the only design difference related to differences in CPS and NHIS reporting patterns (e.g., inequality of beta coefficients for private coverage in reporting about current enrollees in NHIS and CPS).

4.2 Interpretation of Results about Enrollee Characteristics

Findings about the relationships between false-negative reporting and enrollee characteristics suggest that any resulting biases may be similar for surveys using the same reference period. The basis follows:

- Misreporting of enrollees is not random (certain enrollee characteristics are predictive).
- Predictors of misreport tend to be similarly predictive in different survey conditions.
- The overall rate of misreport is similar for enrollees with coverage at the time of the interview.

The nature of the predictive characteristics suggest that level of respondent awareness of coverage may be an underlying factor. Characteristics that plausibly make Medicaid status more salient are associated with less false-negative reporting. These characteristics include:

- Long spell of coverage
- Recent/current spell of coverage
- Recent receipt of medical/prescription service through Medicaid
- Participation in other income-based assistance program

4.3 Interpretation of Results about Design Features

Findings from comparing patterns of false-negative reporting in different survey conditions suggest that asking about current Medicaid status may reduce response error. The basis follows:

- CPS respondents report more accurately for enrollees with coverage at the time of the interview compared to the subset with coverage only during the reference period (controlling for the enrollee characteristics in the model).
- There is little evidence that the problem with reporting non-current enrollees relates to enrollee characteristics, giving credence to the idea that causal factors may be specific to the task of having to recall retrospective enrollment (besides a variable for service in the past 30 days which is expected to differ across these conditions, only variables for TANF and four states have different relationships across CPS conditions).

However, findings also suggest that the effect of recall period and timing of coverage is more complicated than whether or not the respondent must recall retrospective enrollment in order to report the correct

Medicaid status. Similarly, they suggest that CPS is not simply a point-in-time estimate, as some analysts assume. The basis follows:

- NHIS reporting was related to events (receipt of services and duration of coverage) in 2000, which is prior to the NHIS reference period.
- CPS reporting was related to events (receipt of services) in 2001, which is after the CPS reference period.
- But significant effects estimated for events outside the reference period in NHIS and CPS also appear smaller than estimates for the events in the conditions where it was within the reference period.
- Moreover, CPS respondents appear to correctly attend to the retrospective aspect of the reference period for the estimated 27.19% of non-current enrollees who were correctly reported.

Besides reference period, there is reason to suspect that one or more other design differences may have affected CPS and NHIS reporting about enrollees in 2001. The basis follows:

- Tests comparing reporting patterns for the NHIS subset and the CPS-current subset show a difference in the relationships between false-negative report and several enrollee characteristics that have no apparent relationship to reference period (private coverage, SSI, age 18-44, and "other" relationship). These differences should not be attributable to differences in the design of the reference period because there is no opportunity for the causal mechanisms responsible for the reference-period effect to operate (i.e., CPS respondents for current enrollees should derive the same response regardless of whether they report retrospectively or currently). This leaves other differences in design features as likely responsible for the different patterns of response (assuming that the other enrollee characteristics in the model and the survey weights control for differences in the NHIS and CPS samples).
- For example, tests show a different relationship between having private coverage and false-negative reporting in NHIS and CPS current, with private being strongly predictive of misreport in NHIS but not significantly related to misreport in either CPS condition. It may be that the unique use of a flashcard for asking about health insurance in NHIS has something to do with this finding, since the first item on the card is private insurance (and research shows that respondents often mistakenly choose the first item on a list instead of the correct item further down).
- Another example is the difference observed between NHIS and CPS current in their relationships between "other" (relationship to reference person) and misreport. It may be that this has something to do with differences in skip patterns used to administer the surveys, in how the reporting units are defined (multi-family versus single family), and/or in the level of the insurance question (family level versus individual level).

5. Limitations

Specific findings should not be assumed to apply to more recent years of NHIS and CPS.[8] Instead what they suggest about response error in reporting Medicaid should be used to focus on future research on studying the design features that may be most amenable to cost effective modifications. Findings are also limited by the quality of the input files and matching algorithm used to link them.

6. Recommendations for Future Research

In terms of CPS and NHIS specifically, these findings point to the following areas for fruitful research:

- Incorporating paradata to insure that the samples studied here were comparable on mode (CPS and NHIS are nominally face-to-face but some interviews are conducted by telephone) and other aspects of the interview experienced by respondents (such as the length overall and the amount of time before insurance questions).
- The effect of survey topic and the possibility that there may be types of respondents for whom more topical detail does facilitate correct reporting of Medicaid (specifically exploring why a

[8] Beginning with the third quarter in 2004, two new NHIS questions were added to reduce potential errors in reported Medicaid and Medicare status.

labor-focused survey (CPS) appears comparable to a health-focused survey (NHIS) after accounting for their difference in recall period).

More broadly, the findings from this study point to the following areas for fruitful research:

- The sequencing of survey questions and their response options.
- The subject level of questions about Medicaid (e.g., family versus individual level).
- The relationship between salience of Medicaid enrollment and enrollment in other government-assistance programs (and how survey design may enhance the salience of program enrollment more generally).
- The state-specific names surveys use to refer to Medicaid and respondent familiarity with them.

7.Acknowledgements

This paper would not have been possible without the research of the collaborators in the SNACC project. I am especially grateful to people who provided advice specific to this study. Dean Resnick advised in all aspects. Yves Thibaudeau and Paul Allison advised in choice of methodology. Joanne Pascale and Amy O'Hara advised in interpreting and presenting results. Steve Riesz advised in the review of the report.

8. References

Blumberg, S. J., & Cynamon, M. (1999). Misreporting Medicaid Enrollment: Results of Three Studies Linking Telephone Surveys to State Administrative Records. Proceedings of the Seventh Conference on Health Survey Research Methods, DHHS Publication No. (PHS) 01-1013. Hyattsville, MD.

Call, K.T., Davidson, G., Sommers, A.S., Feldman, R., Farseth, P., & Rockwood, T. (2002). Uncovering the Missing Medicaid Cases and Assessing Their Bias for Estimates of the Uninsured. *Inquiry* 38, no. 4: 396–408.

Call, K.T., Davidson, G., Davern, M., Nyman, R. (2008). Medicaid Undercount and Bias to Estimates of Uninsurance: New Estimates and Existing Evidence. *Health Services Research* 43, no. 3:901-14.

Card, D.A., Hildreth, K.G., Shore-Sheppard, L.D. (2001). The Measurement of Medicaid Coverage in the SIPP: Evidence from California, 1990-1996. National Bureau of Economic Research, Working Paper 8514. Cambridge, MA.

Czajka, J. L., Lewis, K. (1999). Using National Survey Data to Analyze Children's Health Insurance Coverage: An Assessment of Issues. Report submitted by Mathematica Policy Research to the U.S. Department of Health and Human Services. Washington, D.C.

Davern, M. E., Call, K. T., Beebe, T. J., Bland, P. , Ziegenfuss, J. & Blewett, L. A. (2008). Validating Health Insurance Coverage Self-Reports: A Comparison Between Self-Reported Coverage and Administrative Data Records. Public Opinion Quarterly Advance Access published on June 1, 2008, DOI 10.1093/poq/nfn013. *Public Opinion Quarterly* 72: 241-259.

Eberly, T., Pohl, M. & Davis, S. (2005). The Maryland Current Population Survey Medicaid Undercount Study. UMBC Center for Health Program Development and Management. Baltimore, MD

Klerman, J. A., Ringel, J.S., Roth, B. (2005). Under-Reporting of Medicaid and Welfare in the Current Population Survey. RAND Labor and Population working paper series WR-169-3. Santa Monica, CA.

Lewis, K., Ellwood, M., & Czajka, J. (1998). Counting the Uninsured: A Review of the Literature, *Assessing the New Federalism*. The Urban Institute Occasional Paper no. 8. Washington, DC. www.urban.org/UploadedPDF/ occ8.pdf (accessed May 2008);

SNACC (The University of Minnesota's State Health Access 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 of the Census). (2008). Examining Discrepancies between the MSIS and the CPS ASEC. www.census.gov/did/www/snacc/SNACC_Phase_2_report.pdf (accessed August 2008).