Misclassification of Medicaid Participation by Dual Eligibles: Evidence from the Medicare Current Beneficiary Survey

Jennifer M. Mellor [Professor],
Economics Director of the Schroeder Center for Health Policy, William & Mary, Williamsburg, VA.

Melissa McInerney [Associate Professor],
Department of Economics, Tufts University, Medford, MA.

Lindsay M. Sabik [Associate Professor]
Department of Health Policy and Management, University of Pittsburgh School of Public Health, Pittsburgh, PA.

Abstract

Previous studies show that survey-based reports of Medicaid participation are measured with error, but no prior study has examined measurement error in an important segment of the Medicaid population – low-income adults enrolled in Medicare. Using the Medicare Current Beneficiary Survey, we examine whether respondent self-reports of Medicaid enrollment match administrative records and present several key findings. First, among low-income Medicare beneficiaries, the false negative rate is 11.5% when the self-report is interpreted as full Medicaid, and 3.7% when the self-report is interpreted as full or partial Medicaid. Second, the likelihood of a false negative report is systematically associated with respondent traits. Third, systematic measurement error results in biased coefficient estimates in models of Medicaid participation defined from self-reports, and the bias is more significant when the researcher interprets self-reports as full Medicaid coverage only. Researchers should use caution when interpreting survey reports as pertaining to full Medicaid coverage only.

Keywords

Introduction

Survey respondents sometimes misreport participation in public programs, causing problems for program officials and researchers (Bitler, Currie, & Scholz, 2003; Bollinger & David, 1997; Colby, Debora, & Heggeness, 2017; Davies & Fisher, 2009; Harris, 2014; Huynh, Rupp, & Sears, 2002; Lynch, 2008; Meijer, Karoly, & Michaud, 2010; Meyer & George, 2011). Some respondents enrolled in a program say they are not (false negative error), while others who are not enrolled say they are (false positive error). In the case of Medicaid, false negative rates can range from 15 to 40% depending on the survey (Boudreaux, Call, Turner,

jmmell@wm.edu (J. Mellor)

Fried, & O'Hara, 2015; Card, Hildreth, & Shore-Sheppard, 2004; Lynch, 2008; Noon, Fernandez, & Porter, 2016), and survey counts of total enrollment can be 10 to 30% less than figures gleaned from administrative data (Card et al., 2004; Klerman, Ringel, & Roth, 2005; Pascale, Roemer, & Resnick, 2009). Further, when measurement error causes a binary dependent variable (like an indicator of Medicaid participation) to be misclassified, it leads to biased coefficient estimates (Bound, Brown, & Mathiowetz, 2001; Hausman, Abrevaya, & Scott-Morton, 1998). The econometrics literature shows that this type of bias depends on the nature of the measurement error, and that, in some cases, the bias can be addressed by using information from validation studies that compare survey reports to administrative data (Meyer & Mittag, 2017).

While several studies examine measurement error in Medicaid participation generally, no prior study has examined measurement error in an important segment of the Medicaid population – low-income adults who are aged or disabled and enrolled in Medicare. Nationwide, about 12 million Medicare beneficiaries receive benefits through Medicaid, including services not covered by Medicare (e.g., long-term services and supports, dental and vision benefits) and financial assistance with Medicare cost-sharing and/or monthly premiums (CMS Medicare-Medicaid Coordination Office, 2017). Given their income, age, and disability status, the average dual eligible has more serious health problems and higher healthcare spending than the typical Medicaid or Medicare beneficiary (CMS Medicare-Medicaid Coordination Office, 2018). As a result, many policy initiatives target the dual-eligible population, such as capitated managed care programs and integrated Medicare-Medicaid financing (Brown & Mann, 2012; Congressional Budget Office, 2013). These initiatives could impact Medicaid take-up, and modeling the impacts of these and other policies using survey data requires a better understanding of the measurement error in survey responses about Medicaid participation by this important population.

Additionally, reports of Medicaid enrollment by potential dual eligibles are likely affected by a type of measurement error that does not apply to other Medicaid-eligible groups. Dual eligibles consist of two groups: full benefit duals who qualify based on the rules in their state Medicaid program, and partial benefit duals who qualify through a Medicare Savings Program (MSP).[i] Both groups receive assistance with premiums and cost-sharing, but only full benefit duals have coverage for benefits that Medicare does not cover (Medicare Payment Advisory Commission, 2008). This full/partial distinction creates another complication when using survey data to study duals' Medicaid participation; this is because none of the major household surveys use coverage questions that distinguish between full and partial Medicaid. This complication is made more important by a significant increase in the number of partial duals in recent years (CMS Medicare-Medicaid Coordination Office, 2017).

---

[i]MSP programs include the Qualified Medicare Beneficiary (QMB), Specified Low-Income Medicare Beneficiary (SLMB), Qualifying Individual (QI), and Qualified Disabled and Working Individuals (QDWI) programs. The programs differ in terms of coverage for out-of-pocket costs and income and asset eligibility criteria.

New Contribution

This study is the first to examine the extent and nature of measurement error in survey reports of Medicaid participation by Medicare enrollees. We use several years of data from the Medicare Current Beneficiary Survey (MCBS), a nationally representative survey that contains both the survey respondent's report of Medicaid coverage plus administrative data on actual Medicaid enrollment and type of coverage (full or partial). First, we describe the extent of measurement error in survey reports of Medicaid participation by the dual-eligible population. Next, we examine how researchers can approach three issues related to measurement error in studies of the determinants of Medicaid participation by duals. We test whether false negative error is systematically associated with respondent traits, which has implications for how the researcher responds to the bias. We test whether the bias is larger when the researcher interprets self-reports of Medicaid participation as reflecting enrollment in either full or partial Medicaid, or full Medicaid only. Lastly, we examine whether measurement error exhibits state-specific differences that vary over time, which has implications for researchers evaluating the effects of state policies on Medicaid take-up using survey data.

Bias in Models with Misclassified Binary Dependent Variables

For context, we briefly review some econometrics on the consequences of measurement error in a binary dependent variable, such as an indicator variable equal to 1 if the respondent has Medicaid coverage, and 0 otherwise. Measurement error in a binary dependent variable, or misclassification, is known to lead to biased estimates of coefficients or marginal effects of the explanatory variables (Bound et al., 2001). Various studies of program participation have used validation data to illustrate the potential for bias in such cases (Bollinger & David, 1997; Davern, Klerman, Baugh, Call, & Greenberg, 2009). An important recent contribution to this literature is Meyer and Mittag (2017), which provides a general assessment of this type of bias and describes how researchers can approach it.

If misclassification is independent of the explanatory variables, the estimated coefficients from a linear regression of the misclassified dependent variable on the explanatory variables (or marginal effects from a probit model) will equal the true coefficients (or marginal effects) multiplied by $(1-\alpha_0-\alpha_1)$, where α_0 is the conditional probability of false positives and α_1 is the conditional probability of false negatives (Bound et al., 2001; Hausman et al., 1998). Thus, when the conditional probabilities sum to less than 1, measurement error in the dependent variable attenuates the estimated coefficients or marginal effects. In this case, the researcher may use the probabilities of false positives and false negatives (assuming they are known) to correct the estimates for the bias, i.e., by dividing the estimates by $(1-\alpha_0-\alpha_1)$. Alternatively, the researcher may safely draw inferences from the relative sizes of the coefficients or marginal effects, since all estimates are attenuated proportionally (Meyer & Mittag, 2017).

If, however, misclassification is correlated with the explanatory variables, the size of the bias and corrections for bias differ. Meyer and Mittag (2017) derive a formula for the bias in linear probability models that can be used to correct coefficient estimates from models of misclassified dependent variables. This correction requires information on the means of the

covariates among false negatives and false positives and the probabilities of false positives and false negatives conditional on the true response. For the probit model, Meyer and Mittag show that additional information from validation data can be used to improve coefficient estimates. In this setting, the researcher could calculate predicted probabilities of false negative and false positive errors with parameter estimates from misclassification models based on validation data (Bollinger & David, 1997). Alternatively, if the researcher has access to validation data, jointly estimating the outcome model and models of false negative and false positive error will improve coefficient estimates over the model using only the misclassified outcome.

In summary, the econometrics literature has established that misclassification of a binary dependent variable results in bias, and that bias can be addressed with different approaches depending on the nature of the measurement error, the model specification, and the availability of validation data. Our analysis provides evidence on these measurement and estimation issues in the context of Medicaid coverage for Medicare enrollees.

Methods

Data

We use data from the 2011–2013 Cost and Use files of the MCBS. Each survey year includes approximately 11,000 Medicare beneficiaries. The MCBS is well-suited to our study for several reasons. First, while multiple household surveys contain information on both Medicare and Medicaid coverage, only the MCBS has survey-based information on Medicaid coverage and information about the respondent's actual Medicaid enrollment from linked administrative records for all respondents.[ii] Second, the MCBS contains rich detail on respondent demographic/economic traits and health status, which allows us to examine the association between measurement error and explanatory variables in models of Medicaid take-up. Third, MCBS data are often used to model Medicaid participation by duals, which helps our analysis inform an existing literature (McInerney, Mellor, & Sabik, 2017; Pezzin & Kasper, 2002; Ungaro & Federman, 2009). We use 2011 for the first year of our study period since the MCBS began using improved sources of administrative data on Medicaid enrollment in that year. Data for 2014 were not released by CMS. We exclude a small number of respondents who were only enrolled in Medicare Part B (n=146 across all three years, less than 0.5% of the overall sample). Since the MCBS employs a rotating panel design, surveying individuals for up to three years, we randomly sampled one observation per respondent, yielding a sample of 19,889 unique Medicare beneficiaries.

The MCBS survey and administrative information on Medicaid coverage are aligned in terms of the reference period. That is, the questionnaire asks respondents whether they had Medicaid coverage for specific months of the year. Similarly, the MCBS indicates specific months of Medicaid coverage obtained from administrative records. However, while the administrative information distinguishes between full and partial Medicaid coverage, the survey does not. This is not unique to the MCBS; insurance coverage questions in other

---

[ii]Restricted data on Medicaid enrollment is available from the HRS, but linkage is not possible for all respondents. Plus, the survey question pertains to coverage over a two-year lookback period, which may increase recall error.

major household surveys do not make this distinction either, as shown in Appendix A. Thus, our analysis of the match between survey response and type of Medicaid coverage may be informative to researchers using a variety of datasets. Appendix B provides more detail on information on Medicaid participation in the MCBS.

Measures

We construct several binary indicators of whether each MCBS respondent had Medicaid coverage during the year. We define one indicator for self-reported Medicaid participation from the survey data, and three from the administrative data (for full Medicaid, partial Medicaid, and either full or partial Medicaid). We then identify whether the survey indicator is misclassified (i.e., survey and administrative coverage indicators do not align); here, we construct two misclassification dummies, allowing for different definitions of the administrative "truth" (i.e., "true" Medicaid coverage is (1) full coverage only or (2) either full or partial coverage). Likewise, we construct two indicators each for false negative error and false positive error.

In addition, we construct measures of respondents' traits including age, sex, race, ethnicity, marital status, education, household income, and state of residence. We also construct several measures of health, including fair or poor self-rated health status, whether the respondent has two or more limitations with activities of daily living (ADLs) or a cognitive impairment, residence in a facility during the year, annual Medicare spending, and the number of months enrolled in Medicaid. Lastly, we construct indicators for the type of Medicare coverage the beneficiary had (Part A only vs. both Parts A and B, and Medicare Advantage vs. traditional Medicare), and whether the respondent had an interview completed by a proxy.

Statistical Analysis

The first part of our analysis describes measurement error in survey responses regarding Medicaid participation by Medicare beneficiaries and compares measurement error across different subgroups. We report rates of self-reported Medicaid participation, misclassification, false negative error, and false positive error for all MCBS respondents, as well as subgroups of low-income respondents and subgroups of important segments of the dual-eligible population (e.g., those above and below age 65, those residing in the community or a facility).

The second part of our analysis examines several aspects of the bias caused by misclassified dependent variables. Since bias corrections depend on whether misclassification error is systematically associated with respondent traits, we first estimate models of false negative error. We estimate linear probability models where the dependent variable equals 1 if the survey measure of Medicaid is a false negative and the explanatory variables are respondent demographic, economic, and health measures. We estimate the models for different samples of known Medicaid enrollees (full, partial, and full/partial combined). This allows us to test whether false negative reports are systematically associated with age, race, income, and other respondent traits, and whether those associations differ between full and partial duals.

Since there is more than one interpretation of the survey response about Medicaid coverage, we next test for differences in the bias that arise under two different researcher interpretations of Medicaid as 1) full Medicaid coverage or 2) either full or partial coverage. We estimate linear probability models of Medicaid participation using alternate definitions of the dependent variable – one in which the dependent variable is the possibly-mismeasured survey response, and another in which the dependent variable is the "true" Medicaid participation measure from the administrative data. We use the estimated coefficients to calculate the size of the bias and examine whether the bias varies under different interpretations of the survey report of Medicaid (i.e., as full Medicaid only, or as either full/ partial Medicaid). In these models, we use explanatory variables used in similar prior studies on elderly Medicaid/SNAP take-up (Borella, De Nardi, & French, 2018; Haider, Jacknowitz, & Schoeni, 2003) as well as characteristics of Medicare coverage. All models also include a full set of state and year dummies.

Lastly, since studies of the effects of state policies on Medicaid take-up rely on the assumption of non-differential time trends across states (Boudreaux et al., 2015), we examine the potential for bias when reporting error exhibits significant state-specific trends. We estimate linear probability models where the dependent variable is either false negative or false positive error, and we test whether there is evidence of state-by-year variation in measurement error.

Results

Descriptive Statistics

Table 1 shows self-reported Medicaid participation rates in the full sample and in various subgroups. Column 2 shows that 23% of all 19,889 respondents in our sample self-report having Medicaid. As expected for a means-tested program, self-reported participation is higher (52%) among low-income Medicare beneficiaries (defined as having household income at or below 150% of the federal poverty guidelines for single and married respondents). Self-reported participation rates are even higher for low-income Medicare beneficiaries residing in facilities (81.9%) and under age 65 (75.7%), which may reflect differences in awareness, need, or eligibility criteria.

Columns 4, 6, and 8 of Table 1 report participation rates among confirmed Medicaid participants (i.e., true positive rates). We report rates separately by persons for whom the administrative record indicates either full or partial coverage, full coverage only, or partial coverage only. The observed true positive rates are high; when "true" Medicaid is defined from an administrative record of either full or partial coverage (column 4), 87.2% of all MCBS respondents on Medicaid and 88.5% of all low-income respondents on Medicaid respond affirmatively on the survey. Further, true positive rates are even higher among full Medicaid enrollees (column 6). More than 95% of all MCBS respondents with full Medicaid and more than 96% of low-income respondents with full Medicaid report enrollment.

Table 1 also shows that many MCBS respondents with only partial Medicaid give survey responses indicating they are on Medicaid (column 8). Nearly 64% of all MCBS respondents (and nearly 66% of low-income respondents) with partial Medicaid report

having Medicaid. These rates are higher among facility residents (89–90%), although these samples are very small. Clearly, many partial duals see themselves as having "Medicaid" coverage; while less than the self-reported participation rates of known full duals (93–98%), the share of known partial enrollees reporting Medicaid coverage is sizeable.

Table 2 extends our descriptive analysis by examining misclassification rates (the shares of all respondents whose survey response does not match the administrative data) as well as false negative and false positive rates among low-income MCBS respondents (i.e., those most likely to be eligible for Medicaid). False negative rates are defined for samples of confirmed Medicaid participants, while false positive rates are defined for confirmed non-participants. We allow for two separate definitions of "true" Medicaid coverage based on administrative records. The top panel defines true Medicaid as either full or partial coverage (a looser definition of Medicaid), and the bottom panel defines true Medicaid as full coverage only (a stricter definition). Several prior Medicaid validation studies focusing on adults of all ages examine full benefit coverage only (Boudreaux et al., 2015; Call, Davern, Klerman, & Lynch, 2013), in one case noting that some survey questions are intended to measure full coverage.

Table 2 shows that misclassification rates are significantly lower under the looser interpretation of true Medicaid. For example, 8.8% of low-income MCBS respondents are misclassified based on survey data when using the looser definition of Medicaid, while 13.4% of low-income MCBS respondents are misclassified under the stricter definition (column 2; p < 0.0001). In columns 3 and 4, we show that the type of misclassification depends on how we define "true Medicaid." Under the looser definition of Medicaid, a larger share of the misclassified responses reflect false negatives, and under the stricter definition of Medicaid, a larger share of the misclassified responses reflect false positives. Accordingly, the false negative rate (defined for those who have Medicaid according to the administrative data) is higher under the looser interpretation of true Medicaid than under the stricter interpretation (e.g., 11.5% versus 3.7% in the low-income sample, as shown in column 5).

Column 6 of Table 2 reports false positive rates, which are defined for respondents who are not enrolled in Medicaid according to the administrative data. False positive rates are higher when we define true Medicaid as full coverage only (the stricter definition). For example, 20.3% of low-income MCBS respondents without full Medicaid say they have Medicaid, while only 5.3% of low-income MCBS respondents without either full or partial Medicaid say they have Medicaid. This difference is consistent with the large numbers of partial duals interpreting their partial coverage as Medicaid.

Is measurement error independent of explanatory variables?

To examine the relationship between misclassification error and respondent traits, we estimate linear probability models of false negative error.[iii] Table 3 presents estimates from separate models for known full Medicaid participants (columns 1–2), known partial

---

[iii]We model false negative errors since that is the type of misclassification more commonly studied in the related prior literature. For interested readers, we report the results from linear probability models of false positive error in Appendix C.

Medicaid participants (columns 3–4) and those known to have either full or partial Medicaid (columns 5–6). In each case, we separately focus on low–income respondents or low-income community-dwelling adults age 65 and up (an oft-studied segment of duals) (Pezzin & Kasper, 2002; Ungaro & Federman, 2009).

The coefficient estimates and F-test results reported in Table 3 provide evidence that false negative reports are associated with respondent traits. False negative error is significantly and positively associated with age, household income, and having a high school or some college education (relative to not completing high school). False negative error is significantly and negatively associated with other race and being unmarried, and with several indicators of poor health: having fair/poor self-rated health, having two or more ADL limitations, having higher Medicare spending, and residing in a facility. That said, demographic and economic traits are more strongly predictive of underreporting of partial Medicaid than underreporting of full Medicaid (as shown by F-statistics of 4.03 and 6.13 in partial Medicaid models compared to only 0.75–1.09 in full Medicaid models). In contrast, health characteristics are more strongly predictive of underreporting of full Medicaid than underreporting of partial Medicaid (as shown by F-statistics of 10.65 and 13.18 in full Medicaid models compared to 1.91 and 4.63 in partial Medicaid models).

We also examine other potential explanations for underreporting using information on proxy interviews, a respondent's time on Medicaid, and the type of Medicare coverage the respondent has. Proxy respondents and persons with more months of Medicaid coverage in the year were less likely to underreport Medicaid, while persons with Medicare Advantage were more likely to underreport Medicaid. We also find evidence of significant differences in underreporting across states, based on tests of the joint significance of state dummy variables in the models (where p-values are 0.001 or lower in five of the six models).

Does the researcher's interpretation of the survey response affect bias?

We next use the MCBS data to calculate the bias arising from using mismeasured survey data to estimate models of Medicaid take-up. Specifically, we investigate a choice available to researchers: if the survey has one self-report of Medicaid coverage, is it better to interpret the survey response as full Medicaid, or as either full or partial Medicaid?

We first calculate the bias from interpreting the survey question as pertaining to the looser definition of an administrative report of either full or partial Medicaid. We use our low-income sample to estimate a linear probability model of Medicaid participation where the dependent variable is equal to 1 if the respondent reported Medicaid coverage on the survey; estimated coefficients are shown in Table 4, column 1. We also estimate a linear probability model where the dependent variable is equal to 1 if the respondent had either full or partial Medicaid according to the administrative records; estimated coefficients are shown in Table 4, column 2. For each coefficient, column 3 reports the bias in the survey as a percentage of the unbiased coefficient (that is, the coefficient estimate from the model of the administrative Medicaid measure, less the coefficient estimate from model of the survey response, divided by the administrative (true) coefficient estimate). We report results from chi-square tests of the hypothesis that the two coefficients are equal to one another (column 4). We then repeat the above steps where the stricter administrative measure of full Medicaid yields the "true"

coefficients; results are reported in columns 1, 5–7 of Table 4. Explanatory variable means are reported in Appendix D.

As expected, the results provide evidence that the survey-based coefficients are biased, sometimes to a sizeable extent. Further, given the non-random nature of the misclassification error, the survey coefficients are not always attenuated. For example, column 3 reports the bias from using survey data to estimate models of full/partial Medicaid participation. The survey data biases the estimated effect of Part A only coverage on full/partial Medicaid take-up so that it is 33% less than the true effect (column 3), but biases the estimated effect of income upward, making it 15.5% larger than the true effect. The former difference is statistically significant. Column 6 reports the bias from using survey data to estimate models of full Medicaid participation; some coefficients are biased upward while others are biased downward, and in several cases, the bias is significantly different from zero.

We next compare the size and impact of the bias in survey data when the administrative measure is the looser definition—reflecting full or partial Medicaid— with the size/impact of the bias under the stricter definition of full Medicaid. A comparison of columns 4 and 7 shows that when interpreting the survey as indicative of full Medicaid coverage, the bias in the survey data yields coefficient estimates that are significantly different from the true coefficients more often than when the researcher interprets the survey as indicative of either full or partial Medicaid coverage. A comparison of columns 3 and 6 shows that, in those instances, the bias is also larger in absolute value when interpreting the survey as indicative of full Medicaid coverage. This suggests there is less potential for bias from misclassification when the survey question is interpreted as either full or partial Medicaid coverage, and not interpreted strictly as full Medicaid.[iv]

Does measurement error in the survey response exhibit state-by-year variation?

Finally, we estimate linear probability models where the dependent variable is either false negative or false positive error to test whether there is evidence of state-by-year variation in measurement error. The results are summarized in Table 5, which reports the results from an F-test of the joint significance of a set of interaction terms between a full set of state dummy variables and both a year 2012 dummy and a year 2013 dummy (full results upon request). The restricted and unrestricted models include state dummies and year dummies, and control for respondent demographic, economic and health traits. We report F-tests from eight models of each outcome, varying the sample and the administrative measure of Medicaid. In most models, the state-by-year coefficient estimates are not jointly significant, but in one of the eight specifications, we are able to reject the null hypothesis that there is no state-by-year variation in survey measurement error of Medicaid (i.e., when estimating false negative reports among low-income individuals 65 and older who reside in the community).[v] It is possible that the small sample sizes contribute to the insignificant F-statistics; however, our

---

[iv]In two robustness checks, we repeat the same steps using (1) low-income community-dwelling respondents age 65 and up, and (2) respondents with incomes 100% FPL. Results (available upon request) exhibit the same pattern that we report in our main results (i.e., using the low-income sample of respondents at or under 150% FPL).
[v]In another investigation of potential for bias in estimates of effects of state-level policy, we estimated models of Medicaid participation separately for survey and administrative report of full Medicaid, including as explanatory variables a dummy equal to one if the state expanded Medicaid, and an interaction of that dummy with year 2012 and year 2013 variables. Estimated coefficients of the

samples sizes are able to identify significant persistent state-specific differences in measurement error. That is, in 10 of 16 models, the state coefficient estimates themselves are jointly significant.

Conclusion

We examine the extent to which survey responses of Medicaid participation are misclassified for individuals also covered by Medicare, a population that has not been examined before in the literature on measurement error in survey data on program participation, and a population of significant policy interest. We highlight four specific findings and their implications for researchers.

First, misclassification errors are considerably less common in our study context – dual eligibles surveyed in the MCBS – than in other settings. We estimate false negative rates of 3.7–11.5% among all low-income MCBS respondents, with the lower value applying to known full Medicaid participants and the higher value pertaining to known full and partial participants. Prior studies report false negative rates of 15% in the 1990–93 Survey of Income and Program Participation (Card et al., 2004), 21.6% in the 2009 ACS (Boudreaux et al., 2015), 30% in the 2001 NHIS (Lynch, 2008), and 40% in the 2006–2011 CPS Annual Social and Economic Supplement (Noon et al., 2016).[vi] Although false positive rates are reported far less often in the literature, our estimates of 5–20% false positive rates in the low-income population are on the low-end of the 20–40% range reported in the ACS (Noon et al., 2016; SNACC Phase V, 2010). As a result, a relatively small fraction of total observations in our various samples are misclassified.

Second, although enumerators and researchers may intend for survey questions to pertain to full Medicaid coverage only, we find that the survey question in the MCBS is picking-up many partial benefit duals. Our results are the first to demonstrate this and do so in several ways: 1) nearly two-thirds of MCBS respondents with confirmed partial Medicaid only (i.e., not full Medicaid) report having Medicaid; 2) the share of misclassified observations is significantly lower when we define "true" Medicaid coverage to be either full or partial coverage than when we use a stricter definition of only full coverage; 3) the bias that results from using survey data is more pronounced when the researcher interprets the survey response as pertaining to full Medicaid, as opposed to either full or partial Medicaid. This suggests that researchers should use caution when interpreting survey data. If full Medicaid coverage is the outcome of interest, researchers may wish to use administrative data in lieu of survey data or consider the bias corrections or robustness checks described in the econometrics literature. In the long run, researchers wishing to use survey data to study full (or partial) coverage exclusively would benefit from the development of separate items on existing survey instruments. Were such items developed, prompts with the names of relevant Medicaid programs could potentially improve response quality.

---

interaction terms did not differ across the two models. We repeated this exercise using the administrative report of partial or full Medicaid. We also repeated this exercise using another state policy variable, a dummy equal to one if the state adopted a Medicaid-Medicare coordination program. In none of these tests did we find any evidence that the "true" coefficients of the state policy-by-year variables differ from the biased coefficients.

[vi]The high false negative rates in the CPS ASEC may be due to imputing procedures or recall difficulties since the question asks about coverage in the prior year.

Third, misclassification error in survey reports of Medicaid is not independent of respondent traits. False negative error varies significantly by respondent traits such as age, race, marital status, and health conditions. The associations are present both for low-income confirmed full duals and duals with either full or partial Medicaid, and they are observed in one of the most commonly-studied segments of the dual population: low-income community-dwelling persons age 65 and up. This suggests that researchers using self-reported data to model Medicaid participation cannot simply report coefficient ratios or scale-up the estimated coefficients from linear probability models. In fact, doing so would exaggerate the bias when misclassification error is not random. Instead researchers may wish to use the corrections described by Meyer and Mittag (2017). As they also show, when validation data or parameter estimates are not available from the same survey, using data obtained from a similar setting may be worthwhile. For that reason we provide some of the necessary data for such corrections in Appendix E. Interested readers can use these data in conjunction with section 2.1 from Meyer and Mittag.

Fourth, the MCBS data suggest that one type of misclassification error—false negative reports—exhibits year-to-year variation that differs across states among low-income community-dwelling older adults. This is notable since many prior studies on the effects of state policies on duals' Medicaid participation use difference-in-difference models which rely on the common trends assumption, or the assumption that pre-period patterns in Medicaid participation are similar in states that experience policy interventions and states that do not (Gardner & Gilleskie, 2012; McInerney et al., 2017; Pezzin & Kasper, 2002; Ungaro & Federman, 2009). While many factors could lead this assumption to be violated, one factor could be the presence of differential trends in misclassification across states; our results suggest that factor may be a concern in some populations, at least in our dataset and time period (2011–2013).

There are limitations to our study. While we add to the literature by using a survey for which the sampling frame is all Medicare enrollees, all of whom have administrative information on coverage, we cannot assess the accuracy and completeness of the administrative data directly. Another limitation is that our results are derived from the MCBS and may not necessarily generalize to other studies. While the design of the question in the MCBS is similar to some other surveys that include specific state Medicaid program names (e.g., NHIS, MEPS, HRS), this is not the case in other surveys (e.g., ACS, CPS, SIPP). That said, the MCBS is used frequently in studies of Medicaid participation, and researchers using MCBS data prior to the 2011 adoption of improved administrative data may find our work especially informative. Finally, we focus only on the consequences of measurement error when Medicaid participation is used as a dependent variable; we note that other econometric considerations arise when mismeasured Medicaid participation is used as an explanatory variable (e.g., Kreider, Pepper, Gundersen, & Jolliffe, 2012; Almada, McCarthy, & Tchernis, 2016; Nguimkeu, Denteh, & Tchernis, 2019).

Dual Medicare-Medicaid enrollees represent a costly and vulnerable population. Research studies on duals – including studies on dual Medicaid enrollment and how it is impacted by policy changes (Borella et al., 2018; Ettner, 1997; Gardner & Gilleskie, 2012; McInerney et al., 2017; Pezzin & Kasper, 2002; Ungaro & Federman, 2009), studies on the association

between dual enrollment and healthcare utilization (Moon & Shin, 2006; Rahman, Tyler, Thomas, Grabowski, & Mor, 2015), or studies of health, disability, and service use within the dual-eligible population (Reichard & Fox, 2013) – can offer important lessons for public policy and public health. Investigations of these issues based on survey data will benefit from careful attention to the potential for misclassification of Medicaid participation and ways to address the bias arising from it.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

References

Almada L McCarthy IM, & Tchernis R (2016). What can we learn about the effects of food stamps on obesity in the presence of misreporting? Amer J of Ag Econ, 98(4), 997–1017

Bitler MP, Currie J, & Scholz JK (2003). WIC eligibility and participation. J Hum Resour, 38(4), 1139–1179.

Bollinger CR, & David MH (1997). Modeling discrete choice with response error: Food stamp participation. J Am Stat Assoc, 92(439), 827–835.

Borella M, De Nardi M, & French E (2018). Who receives Medicaid in old age? Rules and reality. Fisc Stud, 39(1), 65–93. doi:10.1111/1475-5890.12145. [PubMed: 29610542]

Boudreaux MH, Call KT, Turner J, Fried B, & O'Hara B (2015). Measurement error in public health insurance reporting in the American Community Survey: evidence from record linkage. Health Serv Res, 50(6), 1973–1995. [PubMed: 25865628]

Bound J, Brown C, & Mathiowetz N (2001). Measurement error in survey data. In Leamer EE & Heckman JJ (Eds.), Handbook of Econometrics (Vol. 5, pp. 3705–3843).

Brown R, & Mann DR (2012). Best bets for reducing Medicare costs for dual eligible beneficiaries: Assessing the evidence. Retrieved from https://www.kff.org/health-reform/report/best-bets-for-reducing-medicare-costs-for-duals/.

Call KT, Davern ME, Klerman JA, & Lynch V (2013). Comparing errors in m edicaid reporting across surveys: Evidence to date. Health Serv Res, 48(2pt1), 652–664. [PubMed: 22816493]

Card D, Hildreth AKG, & Shore-Sheppard LD (2004). The measurement of Medicaid coverage in the SIPP: Evidence from a comparison of matched records. J Bus Econ Stat, 22(4), 410–420. doi:10.1198/073500104000000208.

CMS Medicare-Medicaid Coordination Office. (2017). Data analysis brief: Medicare-Medicaid dual enrollment 2006 through 2016. Retrieved from https://www.cms.gov/Medicare-Medicaid-Coordination/Medicare-and-Medicaid-Coordination/Medicare-Medicaid-Coordination-Office/DataStatisticalResources/Downloads/Eleven-YearEver-EnrolledTrendsReport_2006-2016.pdf.

CMS Medicare-Medicaid Coordination Office. (2018). People enrolled in Medicare and Medicaid. Retrieved from https://www.cms.gov/Medicare-Medicaid-Coordination/Medicare-and-Medicaid-Coordination/Medicare-Medicaid-Coordination-Office/Downloads/MMCO_Factsheet.pdf.

Colby S, Debora J, & Heggeness ML (2017). How Well Do Individuals Report Supplemental Nutrition Assistance Program (SNAP) Take Up in Household Surveys? SEHSD Working Paper. Retrieved from https://www.census.gov/content/dam/Census/library/working-papers/2017/demo/SEHSD-WP2017-03.pdf.

Congressional Budget Office. (2013). Dual-eligible beneficiaries of Medicare and Medicaid: Characteristics, health care spending, and evolving policies. Retrieved from https://www.cbo.gov/publication/44308.

Davern M, Klerman JA, Baugh DK, Call KT, & Greenberg GD (2009). An examination of the Medicaid undercount in the current population survey: preliminary results from record linking. Health Serv Res, 44(3), 965–987. doi:10.1111/j.1475-6773.2008.00941.x. [PubMed: 19187185]

Davies PS, & Fisher TL (2009). Measurement issues associated with using survey data matched with administrative data from the Social Security Administration. Soc Secur Bull, 69, 1.

Ettner SL (1997). Adverse selection and the purchase of Medigap insurance by the elderly. J Health Econ, 16(5), 543–562. [PubMed: 10175630]

Gardner L, & Gilleskie DB (2012). The effects of state Medicaid policies on the dynamic savings patterns and Medicaid enrollment of the elderly. J Hum Resour, 47(4), 1082–1127.

Haider SJ, Jacknowitz A, & Schoeni RF (2003). Food stamps and the elderly: Why is participation so low? J Hum Resour, 1080–1111.

Harris BC (2014). Within and across county variation in SNAP misreporting: Evidence from linked ACS and administrative records. CARRA Working Paper. Retrieved from https://www.census.gov/library/working-papers/2014/adrm/carra-wp-2014-05.html.

Hausman JA, Abrevaya J, & Scott-Morton FM (1998). Misclassification of the dependent variable in a discrete-response setting. J Econometrics, 87(2), 239–269. doi:Doi 10.1016/S0304-4076(98)00015-3.

Huynh M, Rupp K, & Sears J (2002). The assessment of survey of income and program participation (SIPP) benefit data using longitudinal administrative records. Retrieved from http://www.nber.org/sipp/docs/wp238.pdf.

Klerman J, Ringel J, & Roth B (2005). Under-reporting of Medicaid and welfare in the Current Population Survey. RAND Working Paper. Retrieved from https://www.rand.org/pubs/working_papers/WR169-3.html.

Kreider B, Pepper JV, Gundersen C, & Jolliffe D (2012). Identifying the effects of SNAP (food stamps) on child health outcomes when participation is endogenous and misreported. JASA, 107 (499), 958–975.

Lynch V (2008). Medicaid enrollment: The relationships between survey design, enrollee characteristics, and false-negative reporting. JSM Proceedings, Statistical Computing Section. Alexandria, VA: American Statistical Association. Retrieved from https://ww2.amstat.org/sections/srms/Proceedings/y2008/Files/302382.pdf.

McInerney M, Mellor JM, & Sabik LM (2017). The effects of state Medicaid expansions for working-age adults on senior Medicare beneficiaries. American Economic Journal-Economic Policy, 9(3), 408–438. doi:10.1257/pol.20150402.

Medicare Payment Advisory Commission. (2008). Increasing participation in the Medicare savings programs and the low-income drug subsidy. Retrieved from http://medpac.gov/docs/default-source/reports/Mar08_Ch05.pdf.

Meijer E, Karoly LA, & Michaud P-C (2010). Using matched survey and administrative data to estimate eligibility for the Medicaid Part D Low-Income Subsidy program. Soc Secur Bull, 70, 63. [PubMed: 20560303]

Meyer BD, & George R (2011). Errors in survey reporting and imputation and their effects on estimates of food stamp program participation. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1824261.

Meyer BD, & Mittag N (2017). Misclassification in binary choice models. J Econometrics, 200(2), 295–311. doi:10.1016/j.jeconom.2017.06.012.

Moon S, & Shin J (2006). Health care utilization among Medicare-Medicaid dual eligibles: A count data analysis. BMC Public Health, 6, 88. [PubMed: 16595021]

Nguimkeu P, Denteh A, & Tchernis R (2019). On the estimation of treatment effects with endogenous misreporting. J Econometrics, 208 (2):487–506.

Noon JM, Fernandez L, & Porter SR (2016). Response Error and the Medicaid Undercount in the Current Population Survey. CARRA Working Paper. Retrieved from https://osf.io/preprints/socarxiv/ch8qt/.

Pascale J, Roemer MI, & Resnick DM (2009). Medicaid Underreporting in the CPS Results from a Record Check Study. Public Opin Quart, 73(3), 497–520. doi:10.1093/poq/nfp028.

Pezzin LE, & Kasper JD (2002). Medicaid enrollment among elderly medicare beneficiaries: Individual determinants, effects of state policy, and impact on service use. Health Serv Res, 37(4), 827–847. doi:DOI 10.1034/j.1600-0560.2002.55.x. [PubMed: 12236387]

Rahman M, Tyler D, Thomas KS, Grabowski DC, & Mor V (2015). Higher Medicare SNF care utilization by dual-eligible beneficiaries: Can Medicaid long-term care policies be the answer? Health Serv Res, 50(1), 161–179. [PubMed: 25047831]

Reichard A, & Fox MH (2013). Using population-based data to examine preventive services by disability type among dually eligible (Medicare/Medicaid) adults. Disabil Health J, 6(2), 75–86. doi:10.1016/j.dhjo.2012.12.001. [PubMed: 23507157]

SNACC Phase V (2010). Extending the Phase II analysis of discrepancies between the National Medicaid Statistical Information System (MSIS) and the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC) from calendar years 2000–2001 to calendar years 2002–2005. Retrieved from http://www.shadac.org/publications/snacc-phase-v-report.

Ungaro R, & Federman AD (2009). Restrictiveness of eligibility determination and Medicaid enrollment by low-income seniors. J Aging Soc Policy, 21(4), 338–351. doi:10.1080/08959420903166993. [PubMed: 20092126]

Table 1.

Medicaid Reporting and Underreporting by Type of Respondent and Full/Partial Status

Sample	Total Obs.	Share self-reporting Medicaid	Administrative Record Indicates Full or Partial Medicaid		Administrative Record Indicates Full Medicaid		Administrative Record Indicates Partial Medicaid	
			N	Share self-reporting Medicaid	N	Share self-reporting Medicaid	N	Share self-reporting Medicaid
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All MCBS Respondents	19,889	0.230	4,881	0.872	3,642	0.952	1,239	0.638
Age 65+	16,223	0.150	2,654	0.843	1,957	0.947	697	0.551
Age <65	3,666	0.583	2,227	0.907	1,685	0.958	542	0.749
Community Residents	18,195	0.193	3,819	0.844	2,610	0.943	1,209	0.613
Facility Residents	1,694	0.626	1,062	0.973	1,032	0.975	30	0.900
Age 65+ Community Residents	14,784	0.109	1,829	0.787	1,153	0.931	676	0.540
Low-Income (LI) MCBS Respondents	7,591	0.520	4,265	0.885	3,170	0.963	1,095	0.659
LI, Age 65+	5,104	0.405	2,297	0.854	1,672	0.959	625	0.573
LI, Age <65	2,487	0.757	1,968	0.921	1,498	0.967	470	0.774
LI, Community Residents	6,551	0.473	3,412	0.861	2,343	0.956	1,069	0.654
LI, Facility Residents	1,040	0.819	853	0.980	827	0.983	26	0.885
LI, Age 65+ Community Residents	4,293	0.334	1,662	0.806	1,054	0.946	608	0.564

Notes: Column 1 reports the number of respondents in each of 12 samples constructed from pooled MCBS surveys from 2011–13, and column 2 reports the share of column 1 respondents who report Medicaid coverage on the survey. Column 3, 5, and 7 reports the number of column 1 respondents whose administrative records indicate full/partial Medicaid (col 3) full Medicaid only (col 5), or partial Medicaid only (col 7). Columns 4, 6, and 8 report the share of each of these three groups who report Medicaid coverage on the survey (i.e., the true positive rate for full/partial Medicaid (col 4), the true positive rate for full Medicaid (col 6), and the true positive rate for partial Medicaid (col 8). Low-Income (LI): Income is below 150% of the federal poverty guideline; MCBS: Medicare Current Beneficiary Survey.

Table 2.

Misclassification, False Negative, and False Positive Rates in Samples of Low-Income MCBS Respondents

Sample	"Truth" is Administrative Record of <u>Full or Partial</u> Medicaid Receipt					
			% Misclassified due to:			
	Total Obs.	Share Misclassified	a False Negative Report	a False Positive Report	False Negative Rate	False Positive Rate
	(1)	(2)	(3)	(4)	(5)	(6)
Low-Income (LI) MCBS Respondents	7,591	0.088	74%	26%	0.115	0.053
LI, Age 65+	5,104	0.086	76%	24%	0.146	0.037
LI, Age <65	2,487	0.090	69%	31%	0.079	0.135
LI, Community Residents	6,551	0.097	75%	25%	0.139	0.051
LI, Facility Residents	1,040	0.032	52%	48%	0.020	0.086
LI, Age 65+ Community Residents	4,293	0.096	78%	22%	0.194	0.035

Sample	"Truth" is Administrative Record of <u>Full</u> or Partial Medicaid Receipt					
			% Misclassified due to:			
	Total Obs.	Share Misclassified	a False Negative Report	a False Positive Report	False Negative Rate	False Positive Rate
	(1)	(2)	(3)	(4)	(5)	(6)
Low-Income (LI) MCBS Respondents	7,591	0.134[*]	12%	88%	0.037	0.203
LI, Age 65+	5,104	0.104[*]	13%	87%	0.041	0.135
LI, Age <65	2,487	0.194[*]	10%	90%	0.033	0.439
LI, Community Residents	6,551	0.147[*]	11%	89%	0.044	0.204
LI, Facility Residents	1,040	0.051[*]	26%	74%	0.017	0.183
LI, Age 65+ Community Residents	4,293	0.115[*]	12%	88%	0.054	0.134

Notes: Column 1 reports the number of respondents in each of 6 samples constructed from pooled MCBS surveys from 2011–13. Column 2 reports the share of column 1 respondents whose survey response does not match the administrative record, where the administrative record is defined as either full/partial Medicaid (top panel), or full Medicaid (bottom panel). Columns 3 and 4 report the percent of all the misclassified (unmatched) observations arising from either false negative reports (i.e., responding that one does not have Medicaid when the administrative record says otherwise) or false positive reports (i.e., responding that one has Medicaid when the administrative record says otherwise). Column 5 reports the false negative rate, defined as the number of respondents reporting that they do not have Medicaid, divided by the total number of respondents with Medicaid according to the administrative record. Column 6 reports the false positive rate, defined as the number of respondents reporting that they have Medicaid, divided by the total number of respondents without Medicaid according to the administrative record. Low-Income (LI): Income is below 150% of the federal poverty guideline; MCBS: Medicare Current Beneficiary Survey.

[*] indicates that the misclassification rate in the bottom panel is significantly different from that in the top panel at the .01 level or lower.

Table 3.

Linear Probability Models of False Negative Responses (Dep Var = 1 if Survey Response is Not on Medicaid)

Explanatory Variable	Respondents with Full Medicaid from Administrative Record		Respondents with Partial Medicaid from Administrative Record		Respondents with Full or Partial Medicaid from Administrative Record	
	Low Income Respondents	LI, Age 65+ Community Residents	Low Income Respondents	LI, Age 65+ Community Residents	Low Income Respondents	LI, Age 65+ Community Residents
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.004	0.005	0.009	0.005	−0.000	0.014
	(0.008)	(0.016)	(0.034)	(0.053)	(0.013)	(0.023)
Age >=75 & <85	0.027**	0.023	0.098**	0.056	0.074***	0.045**
	(0.011)	(0.015)	(0.039)	(0.047)	(0.016)	(0.022)
Age >= 85	0.021	0.026	0.099*	0.080	0.058***	0.052*
	(0.015)	(0.020)	(0.055)	(0.064)	(0.022)	(0.029)
Black	0.003	−0.003	0.019	0.026	−0.005	−0.010
	(0.010)	(0.018)	(0.039)	(0.056)	(0.015)	(0.026)
Asian	−0.002	−0.033	0.080	0.114	−0.030	−0.063
	(0.020)	(0.027)	(0.105)	(0.141)	(0.032)	(0.045)
Other race	−0.009	−0.029	−0.087	−0.064	−0.044**	−0.059*
	(0.013)	(0.022)	(0.061)	(0.089)	(0.021)	(0.035)
Hispanic	0.008	−0.016	−0.011	−0.014	−0.026	−0.055*
	(0.012)	(0.019)	(0.052)	(0.071)	(0.018)	(0.029)
Widowed	−0.009	−0.012	0.008	−0.040	−0.001	−0.040
	(0.013)	(0.020)	(0.048)	(0.063)	(0.020)	(0.029)
Separated/divorced	−0.017	−0.026	−0.104**	−0.180***	−0.061***	−0.110***
	(0.013)	(0.021)	(0.044)	(0.064)	(0.019)	(0.030)
Never married	−0.006	0.010	−0.145***	−0.280**	−0.093***	−0.116***
	(0.013)	(0.027)	(0.049)	(0.113)	(0.019)	(0.042)
Income as % FPL	0.023	0.006	0.298***	0.336***	0.224***	0.287***
	(0.015)	(0.026)	(0.059)	(0.081)	(0.022)	(0.036)
High school graduate	0.003	0.010	0.084**	0.151***	0.034**	0.095***
	(0.009)	(0.017)	(0.036)	(0.051)	(0.014)	(0.025)
Some college	−0.003	0.004	0.008	0.091	0.023	0.049
	(0.011)	(0.022)	(0.039)	(0.062)	(0.016)	(0.030)
College grad or higher	0.013	0.025	−0.171**	−0.205**	−0.026	−0.042
	(0.019)	(0.031)	(0.075)	(0.095)	(0.029)	(0.044)
Fair/poor self-rated health	−0.009	−0.020	−0.007	0.033	−0.020*	−0.026
	(0.008)	(0.014)	(0.032)	(0.049)	(0.012)	(0.021)

Explanatory Variable	Respondents with Full Medicaid from Administrative Record		Respondents with Partial Medicaid from Administrative Record		Respondents with Full or Partial Medicaid from Administrative Record	
	Low Income Respondents	LI, Age 65+ Community Residents	Low Income Respondents	LI, Age 65+ Community Residents	Low Income Respondents	LI, Age 65+ Community Residents
	(1)	(2)	(3)	(4)	(5)	(6)
Two or more ADLs	−0.001	0.002	−0.056	−0.074	−0.040***	−0.053**
	(0.009)	(0.015)	(0.035)	(0.051)	(0.013)	(0.022)
Dementia or Alzheimer's	−0.007	0.006	0.085	0.081	−0.006	−0.006
	(0.013)	(0.019)	(0.061)	(0.084)	(0.021)	(0.031)
Log (Medicare spending), 2013 $	−0.004**	−0.007**	−0.014**	−0.012	−0.013***	−0.019***
	(0.002)	(0.004)	(0.007)	(0.010)	(0.003)	(0.005)
Months covered by Medicaid	−0.015***	−0.019***	−0.020***	−0.018**	−0.019***	−0.021***
	(0.002)	(0.003)	(0.006)	(0.008)	(0.002)	(0.004)
Medicare Part A only	−0.061	−0.032	0.618		−0.027	−0.091
	(0.058)	(0.089)	(0.444)		(0.100)	(0.165)
Medicare Advantage	−0.016	−0.020	0.071**	0.054	0.046***	0.050**
	(0.010)	(0.017)	(0.033)	(0.045)	(0.014)	(0.023)
Facility resident	−0.005		−0.253*		−0.091***	
	(0.018)		(0.140)		(0.030)	
Proxy interview	−0.016*	−0.050***	−0.047	−0.081	−0.054***	−0.088***
	(0.009)	(0.016)	(0.043)	(0.064)	(0.014)	(0.025)
Year 2012	−0.009	−0.019	−0.004	−0.002	−0.010	−0.012
	(0.010)	(0.016)	(0.036)	(0.052)	(0.014)	(0.024)
Year 2013	0.002	−0.012	−0.020	−0.046	−0.006	−0.030
	(0.009)	(0.015)	(0.034)	(0.049)	(0.013)	(0.022)
Constant	0.211***	0.321***	0.293**	0.251	0.302***	0.363***
	(0.035)	(0.064)	(0.120)	(0.171)	(0.051)	(0.087)
F test, demographic/economic controls	1.09 (0.364)	0.75 (0.719)	6.13 (0.000)	4.03 (0.000)	16.7 (0.000)	9.51 (0.000)
F test, health controls	13.18 (0.000)	10.65 (0.000)	4.63 (0.000)	1.91 (0.092)	17.97 (0.000)	10.89 (0.000)
F test, year controls	0.76 (0.467)	0.72 (0.485)	0.20 (0.823)	0.55 (0.580)	0.25 (0.779)	0.94 (0.392)
F test, state fixed effects	3.72 (0.000)	2.54 (0.000)	2.24 (0.000)	1.73 (0.000)	1.85 (0.001)	1.07 (0.362)
Observations	2,145	955	960	561	3,105	1,516
Dependent Mean	0.033	0.044	0.343	0.431	0.129	0.187
R-squared	0.115	0.174	0.213	0.222	0.144	0.176

Notes: Standard errors are shown in parentheses below coefficient estimates; statistical significance indicated by

***
(p<0.01),

**
(p<0.05), and

*
(p<0.1).

Models also include a full set of state indicator variables (coefficients not shown). P-values are reported in parentheses below F-tests.

Table 4.

Linear Probability Models of Medicaid Participation and Estimated Bias from Survey Data Misclassification, Low Income Respondent Sample

Explanatory Variable	Dep Var is Survey Report of Medicaid	Dep Var is Admin. Report of Full or Part. Medicaid	Bias in survey data	p-value from χ^2 test of $\beta_S=\beta_A$	Dep Var is Admin. Report of Full Medicaid Only	Bias in survey data	p-value from χ^2 test of $\beta_S=\beta_A$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	0.013	0.024[**]	45.83%	0.493	0.042[***]	69.05%	0.054
	(0.011)	(0.011)			(0.010)		
Age >=75 & <85	−0.140[***]	−0.129[***]	−8.53%	0.567	−0.089[***]	−57.30%	0.005
	(0.013)	(0.013)			(0.013)		
Age >= 85	−0.193[***]	−0.200[***]	3.50%	0.756	−0.130[***]	−48.46%	0.004
	(0.016)	(0.016)			(0.016)		
Black	0.123[***]	0.106[***]	−16.04%	0.378	0.079[***]	−55.70%	0.025
	(0.014)	(0.014)			(0.013)		
Asian	0.215[***]	0.253[***]	15.02%	0.416	0.222[***]	3.15%	0.890
	(0.031)	(0.031)			(0.030)		
Other race	0.105[***]	0.088[***]	−19.32%	0.553	0.080[***]	−31.25%	0.406
	(0.020)	(0.020)			(0.020)		
Hispanic	0.133[***]	0.124[***]	−7.26%	0.723	0.118[***]	−12.71%	0.534
	(0.017)	(0.017)			(0.017)		
Widowed	0.075[***]	0.091[***]	17.58%	0.424	0.040[***]	−87.50%	0.079
	(0.015)	(0.015)			(0.014)		
Sep./divorced	0.220[***]	0.239[***]	7.95%	0.361	0.161[***]	−36.65%	0.004
	(0.015)	(0.015)			(0.014)		
Never married	0.321[***]	0.341[***]	5.87%	0.368	0.324[***]	0.93%	0.892
	(0.015)	(0.015)			(0.015)		
Income as % FPL	−0.253[***]	−0.219[***]	−15.53%	0.146	−0.246[***]	−2.85%	0.752
	(0.015)	(0.015)			(0.015)		
HS graduate	−0.082[***]	−0.087[***]	5.75%	0.759	−0.063[***]	−30.16%	0.248
	(0.012)	(0.012)			(0.011)		
Some college	−0.132[***]	−0.134[***]	1.49%	0.889	−0.151[***]	12.58%	0.293
	(0.014)	(0.014)			(0.013)		
College or higher	−0.182[***]	−0.197[***]	7.61%	0.634	−0.159[***]	−14.47%	0.457
	(0.021)	(0.021)			(0.021)		
Fair/poor health	0.062[***]	0.062[***]	0.00%	0.978	0.057[***]	−8.77%	0.744

Explanatory Variable	Dep Var is Survey Report of Medicaid	Dep Var is Admin. Report of Full or Part. Medicaid	Bias in survey data	p-value from χ² test of βS=βA	Dep Var is Admin. Report of Full Medicaid Only	Bias in survey data	p-value from χ² test of βS=βA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	(0.011)	(0.011)			(0.010)		
Two or more ADL	0.091***	0.083***	−9.64%	0.654	0.106***	14.15%	0.368
	(0.012)	(0.012)			(0.011)		
Dementia or Alzheimer's	0.068***	0.054***	−25.93%	0.562	0.063***	−7.94%	0.835
	(0.017)	(0.017)			(0.016)		
log of Medicare spending, 2013 $	0.026***	0.029***	10.34%	0.427	0.021***	−23.81%	0.087
	(0.002)	(0.002)			(0.002)		
Medicare Part A only	−0.174***	−0.260***	33.08%	0.038	−0.144***	−20.83%	0.476
	(0.034)	(0.034)			(0.033)		
Medicare Advantage	−0.049***	−0.024**	−104.17%	0.154	−0.066***	25.76%	0.312
	(0.012)	(0.012)			(0.012)		
Facility resident	0.248***	0.212***	−16.98%	0.109	0.303***	18.15%	0.021
	(0.017)	(0.017)			(0.017)		
Year 2012	−0.009	−0.011	18.18%	0.890	−0.008	−12.50%	0.954
	(0.012)	(0.012)			(0.012)		
Year 2013	−0.013	−0.023**	43.48%	0.522	−0.017	23.53%	0.800
	(0.011)	(0.011)			(0.011)		
Constant	0.408***	0.411***			0.223***		
	(0.035)	(0.035)			(0.034)		
Observations	7,263	7,263			7,263		
Dependent Mean	0.512	0.554			0.406		
R-squared	0.331	0.323			0.355		

Notes: Standard errors in parentheses; statistical significance indicated by

***
(p<0.01),

**
(p<0.05), and

*
(p<0.1).

Models also include a full set of state indicator variables (coefficients not shown). Columns 1, 2, and 5 report the estimated coefficients from linear probability models of Medicaid participation, where the dependent variable is defined from either survey reports (col 1), administrative records of full or partial Medicaid (col 2), or administrative records of full Medicaid (col 3). The bias estimates presented in columns 3 and 6 are computed as $\{(\beta_A - \beta_S)/\beta_A\}*100$ where β_S is the corresponding coefficient estimate from column 1; the bias presented in column 3 uses the coefficient estimates in column 2 as β_A, and the bias presented in column 6 uses the coefficient estimates in column 5 as β_A. Columns 4 and 7 report the p-values from chi-squared tests that the biased and unbiased coefficients are equal. The p-value in column 4 is from the χ^2 test of equality of the coefficients in columns 1 and 2; the p-value in column 7 is from the χ^2 test of equality of the coefficients in columns 1 and 5.

Table 5.

Tests of State-by-Year Variation in Survey Measurement Error

Sample	Administrative Measure of Medicaid	Dependent Variable	
		Equal to 1 if false negative, 0 if true positive	Equal to 1 if false positive, 0 if true negative
Explanatory variables <u>include</u> proxy interview status			
Low Income Respondents	Equal to 1 if full or partial Medicaid	F=1.02 P-value = 0.4233 n=3,105	F=0.83 P-value = 0.8357 n=2,723
Low Income Respondents	Equal to 1 if full Medicaid only	F=0.74 P-value = 0.9459 n=2,145	F=0.63 P-value = 0.9937 n=3,683
LI, Age 65+ Community Residents	Equal to 1 if full or partial Medicaid	F=1.07 P-value = 0.3196 n=1,516	F=0.83 P-value = 0.8391 n=2,295
LI, Age 65+ Community Residents	Equal to 1 if full Medicaid only	F=1.35 P-value = 0.0440 n=955	F=0.63 P-value = 0.9937 n=2,856
Explanatory variables <u>exclude</u> proxy interview status			
Low Income Respondents	Equal to 1 if full or partial Medicaid	F=1.02 P-value = 0.4221 n=4,024	F=0.93 P-value = 0.6401 n=3,239
Low Income Respondents	Equal to 1 if full Medicaid only	F=0.95 P-value = 0.6053 n=2,951	F=0.57 P-value = 0.9987 n=4,312
LI, Age 65+ Community Residents	Equal to 1 if full or partial Medicaid	F=1.21 P-value = 0.1146 n=1,619	F=0.89 P-value = 0.7260 n=2,595
LI, Age 65+ Community Residents	Equal to 1 if full Medicaid only	F=1.38 P-value = 0.0321 n=1,019	F=0.61 P-value = 0.9958 n=3,195

Notes: All models include the same controls shown in Table 3 (including state fixed effects) plus a full set of state dummies interacted with each year dummy. The F tests report the results from the null hypothesis that the state-by-year interaction term coefficients are jointly equal to zero. We exclude proxy interview status in the bottom panel to increase sample size. Full model results are available upon request.