

# WORKING P A P E R

---

## Under-Reporting of Medicaid and Welfare in the Current Population Survey

JACOB ALEX KLERMAN, JEANNE S. RINGEL, AND  
BETH ROTH

WR-169-3

March 2005

This product is part of the RAND Labor and Population working paper series. RAND working papers are intended to share researchers' latest findings and to solicit additional peer review. This paper has been peer reviewed but not edited. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.



LABOR AND POPULATION

# **Under-Reporting of Medicaid and Welfare in the Current Population Survey**

Jacob Alex Klerman, Jeanne S. Ringel, and Beth Roth

Latest Revision: March, 2005

Address Correspondence to: Jacob Alex Klerman, RAND, 1700 Main St., Santa Monica, CA 90407, 1-310-393-0411, x289, [Jacob\\_Klerman@rand.org](mailto:Jacob_Klerman@rand.org).

This research was jointly funded by the California Health Care Foundation and the U.S. Department of Health and Human Services, Administration for Children and Families (Grant No. 01ASPE377A). The U.S. Bureau of the Census and the California Department of Health Services provided the data. The research in this paper was conducted while the authors were Research Associates at the Center for Economic Studies, U.S. Census Bureau. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the U.S. Census Bureau, the Center for Economic Studies, the California Department of Health Services, RAND, or its sponsors including U.S. Department of Health and Human Services, Administration for Children and Families and the California Health Care Foundation.

This research has benefited greatly from the comments of Leonard Sternbach, and Richard Bavier at US DHHS; Ingrid Aguirre-Happoldt and Chris Perrone at the California Health Care Foundation; participants at a session at the Population Association of America Annual Meetings; and an anonymous RAND reviewer.

## Summary

High-quality survey data are crucial to our understanding of the effects of the Medi-Cal program in California, and the nation's social welfare system more broadly. We can tabulate the number of people enrolled in Medi-Cal from the official program records, the Medi-Cal Eligibility Data System (MEDS). However, beyond enrollment counts, understanding Medi-Cal's effects often requires survey data because information is needed on both enrollees and non-enrollees. For example, to assess take-up rates we need to know the number of people enrolled as well as the number of people who are eligible for the program. If we want to look at take-up by sub-group, we need more detailed information about the characteristics (e.g., family structure, household income) of enrollees and non-enrollees. If we are interested in assessing overall levels of health insurance coverage, we need information on the full population (enrollees and non-enrollees) and their private health insurance coverage. Addressing policy questions of this form requires survey data.

### **The Current Population Survey (CPS), Under-Reporting, and Matching**

The U.S. Bureau of the Census's March Annual Demographic Survey (ADS) to the Current Population Survey (CPS) is the standard data source for analyses of the Medi-Cal program and the nation's social welfare system more broadly. The CPS is a large (about 50,000 households nationally, 6,000 households in California), household survey with information on program participation (including Medicaid/Medi-Cal and welfare), health insurance coverage, and other household characteristics. Two other features of the CPS data are crucial for policy analyses: (1) The ADS data are collected annually in a relatively consistent manner back to the late 1980s—allowing trend and time series analyses; and (2) The data are released promptly—results of the interviews conducted in March are publicly released in late-August or early-September of the same year—allowing nearly real-time tracking of changes.

Unfortunately, the CPS is known to under-report program participation, including Medi-Cal. The official CPS report notes the problem explicitly:

The Current Population Survey (CPS) underreports medicare [stet] and medicaid [stet] coverage compared with enrollment and participation rates from the Centers for Medicare and Medicaid Services (CMS), formerly the Health Care Financing Administration. A major reason for the lower CPS estimates is that the CPS is not designed to collect health insurance data; instead, it is largely a labor force survey. Consequently, interviewers receive less training on health insurance concepts. Additionally, many people may not be aware that they or their children are covered by a health insurance program if they have not used covered services recently and therefore fail to report coverage. CMS data, on the other hand, represent the actual number of people (who) enrolled or participated in these programs and are a more accurate source of coverage levels.

Furthermore, some analyses suggest that the problem has gotten worse over time.

As we will show below, the under-reporting is substantial, but neither its causes, nor its effects, are well understood. Therefore, with funding from the Medi-Cal Policy Institute and the U.S. Department of Health and Human Services, Administration for Children and Families and the cooperation of the U.S.

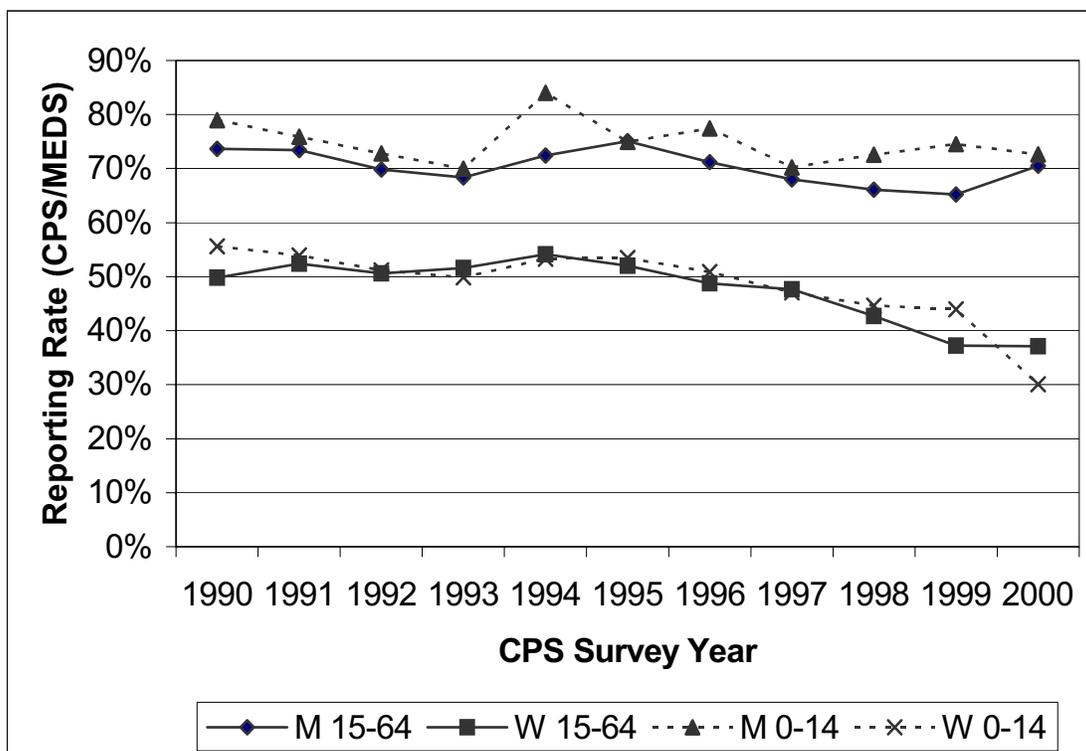
Bureau of the Census and the California Department of Health Services (CDHS), we matched individual-level CPS responses to their corresponding MEDS administrative data records. Specifically, as part of its interview, the CPS attempts to collect Social Security Numbers (SSNs) on all respondents age 15 and older. The MEDS data include SSNs for each enrollee. For this project, the Census Bureau supplied a version of the CPS data for 1990 to 2000 that included a scrambled version of the SSN, where available. In addition, the Census Bureau processed a version of the MEDS data for 1989 to 2001 replacing the original SSNs with the same scrambled SSNs. Where possible, we then matched the two files creating a single analysis file with both CPS and MEDS data. To preserve the confidentiality of CPS respondents and Medi-Cal enrollees, the data analysis took place at the UCLA site of the Secure Data Facility of the Census Bureau's California Census Research Data Center. The authors had no access to identifiers (names or Social Security Numbers) and all research results were reviewed to assure that they did not indirectly reveal the identity of or information about CPS respondents or Medi-Cal enrollees.

## The Magnitude of Under-Reporting And Our Imputation Model

How serious is the problem of under-reporting? Previous analyses of this question using unmatched data have been limited by the inconsistencies between the two data sources. The CPS, administered in March, asks about program enrollment *at any time* in the last calendar year (i.e., the 2000 CPS asks about program participation between January and December 1999). Aggregate Medi-Cal data is usually reported in terms of persons covered per month. The extent to which discrepancies in aggregate counts based on unmatched data were real as opposed to being an artifact of different data concepts has therefore been unclear. Given the structure of our matched data, we can tabulate the individual level Medi-Cal data from MEDS to be consistent with the CPS questions and thus provide a better estimate of under-reporting in the CPS.

Figure S.1 summarizes that analysis. It considers two age groups (adults—15-65 at the interview, and children—0-14 at the interview) and two program concepts: all Medi-Cal (M) and cash assistance/welfare (W—Welfare). Averaged over the entire period, CPS estimates of total Medi-Cal enrollment for adults are only 70 percent of the counts from the official MEDS administrative data, i.e., Medi-Cal is under-reported by about 30 percent. For children, reporting of Medi-Cal is slightly better, about 75 percent. Unlike some national estimates, there is little evidence of a decline in reporting over time.

This overall pattern in Medi-Cal hides a strong divergence by Medi-Cal sub-program. Enrollment in welfare is severely under-reported. Over the entire time period CPS estimates of total welfare enrollment are only 48 percent of the counts from the official MEDS administrative data. For children, the corresponding figure is 51 percent. For welfare, there is clear evidence of a sharp drop in reporting rates over time. The timing of the drop (in the late 1990s) is nearly simultaneous with the implementation of welfare reform in California (i.e., CalWORKs), perhaps suggesting an increase in the stigma of welfare participation.



**Figure S.1 – Reporting Rates (CPS relative to MEDS) by Age and Program**

Source: Tabulations from RAND Merged MEDS File

The under-reporting of program participation in the CPS is severe enough to have substantively important effects on our understanding of the effects of the Medi-Cal program. In this report, we consider two effects. First, under-reporting will lead us to under-estimate take-up rates (the fraction of eligibles enrolled in the program) and thus to over-estimate the need for efforts to increase enrollment or new programs to provide additional coverage. Second, under-reporting will lead us to over-estimate the total number of uninsured people.

Our analysis proceeds as follows. For adults providing an SSN, we overwrite the CPS Medi-Cal responses with the official information from the Medi-Cal administrative data (i.e., treating the MEDS information as the truth). Following Census practice, for children whose parents provide an SSN, we impute Medi-Cal if either the survey response implies Medi-Cal for the child or the administrative data for the parent implies Medi-Cal for the mother (or if not mother, the household head)—in which case the child is almost always covered. However, our ability to match the survey and administrative data is constrained by the fact that only about 62 percent of CPS adults provide an SSN. Furthermore, children under 15 were never asked for an SSN. To address this problem, we build an imputation model to predict mis-reporting among those people without an SSN who we cannot match to the MEDS data. The response errors (i.e., reporting no Medi-Cal in the CPS given actually having Medi-Cal and reporting having Medi-Cal in the CPS given not actually having Medi-Cal) among those not providing an SSN are assumed to follow the general pattern in the sub-sample who do provide an SSN, with an adjustment to force the totals to align exactly (see the full report for details). The problem is more pronounced for children since SSNs are not collected in the CPS for people under age 15. To address this issue we use a combination of information from the head of household and our imputation model. Specifically, where

the mother (or the head of the household if the mother is not present) provides an SSN (as is true for about 66 percent of CPS children), we use the mother or head's Medi-Cal status (from the MEDS or from our imputation) to impute Medi-Cal status to the child. Some Medi-Cal programs include children, but not adults. Therefore, in cases where the child has Medi-Cal, but the head of household does not, the child's data are not changed. Again, as with adults, the imputation includes an adjustment to force the CPS totals (after imputation) to align exactly to the MEDS counts (again, see the full report for details).

These imputations are performed for every observation in the CPS. The resulting individual level file allows us to construct improved estimates of take-up rates and uninsurance coverage. Using the individual-level imputation file, we can consider the effects of under-reporting by respondent characteristics (e.g., gender, age, income).

## **The Effects of Under-Reporting on Estimates of Medi-Cal Enrollment Rates**

If Medi-Cal enrollment is under-reported, then Medi-Cal enrollment rates—the fraction of a demographic group enrolled in Medi-Cal—will also be under-reported. (We note that these are not standard take-up rates that attempt to model actual eligibility from the survey data.) Our analyses of the matched file suggest that the under-reporting is not uniform across sub-groups of the population, so the effects of under-reporting on take-up rates are also not uniform.

Overall Medi-Cal enrollment increases by about 40 percent when we adjust for under-reporting using our imputation model. The increases are slightly larger for adults (42 percent) and slightly smaller for children (38 percent). Consistent with an explanation of under-reporting due to stigma, the increases are smallest for single women with children who are in poverty, largest for those between one and two times poverty, and large for those at more than twice poverty.

Consistent with the even more severe under-reporting, compared to Medi-Cal, the levels of welfare enrollment are lower and the adjustments have a larger effect. The average adjustment more than doubles enrollment rates. The adjustments are similar across children and adults and are smaller for those near poverty, and larger for those out of poverty.

## **The Effects of Under-Reporting on Estimates of Uninsurance**

Another consequence of under-reporting of Medi-Cal enrollment is that it will lead to over-estimates of the rates of uninsurance in the CPS. The magnitude of the over-estimate will depend on the extent to which those under-reporting have other sources of health insurance at some point during the year. If it were the case that everyone who under-reports Medi-Cal did not have any other source of insurance, then we could construct a better estimate of the number of uninsured by subtracting the estimate of under-reporting (i.e., the percent of people in the CPS who report no Medi-Cal, but who our imputation model, based on the matched data, suggests are enrolled) from the raw estimate of the percent of people who are uninsured in the CPS. Conversely, if it were the case that everyone who under-reports Medi-Cal also has private health insurance, then under-reporting would have no effect on the estimates of the uninsured. Our analyses suggest that the truth lies somewhere between these two extremes. Plausibly, we find that under-reporting is more common among those with private health insurance, but under-reporting also includes large numbers of people without private health insurance.

From our matched file, we tabulate rates of other health insurance among people who under-report Medi-Cal. Here we report adjusted estimates of uninsurance based on several different scenarios. We estimate that under-reporting is about 4.1 percent for adults (i.e., 4.1 percent have Medi-Cal but do not report it to the CPS). Consistent with much higher rates of Medi-Cal coverage for children, the corresponding rate of under-reporting is much higher. We estimate that 9.0 percent of all children have Medi-Cal, but do not report it. The unadjusted, or raw, estimate of uninsurance, is 23.5 percent for adults; for children, the unadjusted estimate is slightly lower, 17.8 percent.

The question is: How to use the information from the matched data on the rate of under-reporting to adjust the survey data. A naive approach would, implicitly assume no dual coverage, and subtract the under-reporting from the unadjusted survey estimate of uninsurance. In fact, among those who report in the survey that they have Medi-Cal about a fifth (23.9 percent for adults, 16.7 percent for children) also report private health insurance. Simple tabulations of the matched data suggest that people who have Medi-Cal (according to the survey data), but report that they do not have Medi-Cal in the survey data are much more likely to be dual-covered (i.e., to have other health insurance): 32.3 percent for adults, 34.5 percent for children. Thus, the effect of under-reporting on uninsurance rates is considerable smaller than would be implied by simple subtraction. Using the full imputation model, we estimate uninsurance rates for adults of 20.8 percent (vs. the simple estimate of 23.5 percent) and 11.9 percent for children (vs. 17.8 percent).

Again, adjustments for under-reporting and dual-coverage are not uniform across sub-groups. Sub-groups with higher Medi-Cal receipt rates have larger percentage increases in Medi-Cal coverage with imputation. For adults, the differences across sub-groups are large. For children, the differences across sub-groups are not large.

## Summary

We considered the quality of Medi-Cal information in the Current Population Survey, the standard data source for tabulations of Medi-Cal take-up and levels of uninsurance. The analyses are based on an imputation model derived from a match of individual-level survey data with individual-level administrative data for the Medi-Cal program. We find sizable under-reporting of Medi-Cal, leading to sizable under-estimates of Medi-Cal take-up and sizable over-estimates of the fraction of Californians who are uninsured. These results cover the period 1990 to 2000. The Census Bureau made some adjustments to the CPS interview towards the end of this period. Nevertheless, these results suggest caution in basing policy on unadjusted analyses of the CPS data. Analyses based on unadjusted data are likely to substantially overestimate the magnitude of the problem, especially for children.