

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

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

Latest Revision: February, 2005

Address Correspondence to: Jacob Alex Klerman, RAND, 1700 Main St., Santa Monica, CA 90407, 1-310-393-0411, x289, 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.

Preface

Conventional estimates of the number of uninsured Californians are derived from the Current Population Survey (CPS). Unfortunately, CPS estimates of the number of people receiving Medi-Cal and welfare (AFDC/CalWORKs) are well below the numbers implied by official Medi-Cal records, suggesting that the conventional estimates of the number of uninsured Californians (and their characteristics) are seriously flawed.

To improve our understanding of these issues, the California HealthCare Foundation (through its then separate the Medi-Cal Policy Institute—MCPI) and the U.S. Department of Health and Human Services, Administration for Children and Families (DHHS-ACF) funded RAND to match CPS data to individual-level administrative data for the Medi-Cal program. With the cooperation of the California Department of Health Services (CDHS), the U.S. Bureau of the Census, and the California Census Research Data Center (CCRDC), that match was performed. This document describes the findings of the analysis of those matched data.

Table of Contents

Preface	iii
Table of Contents	v
List of Tables	vii
List of Figures.....	ix
Summary.....	xi
The Current Population Survey (CPS), Under-Reporting, and Matching.....	xi
The Magnitude of Under-Reporting And Our Imputation Model.....	xii
The Effects of Under-Reporting on Estimates of Medi-Cal Enrollment Rates	xiv
The Effects of Under-Reporting on Estimates of Uninsurance	xiv
Summary.....	xv
Acknowledgments.....	xvii
Glossary, List of Symbols, etc.	xix
1. Introduction.....	1
Plan of the Report	1
Previous Literature on Under-Reporting	2
2. Medi-Cal, the CPS, The MEDS, and Under-Reporting.....	5
The Medi-Cal Program	5
The Medi-Cal Eligibility Data System (MEDS)	6
The Current Population Survey (CPS).....	8
Levels and Trends In Mis-Reporting	11
Discussion.....	12
3. The Matched Data.....	13
Matching	13
Simple Reporting Rates.....	17
Who Mis-reports?	19
“Behavioral Regressions” and “Imputational Regressions”	21
Understanding the Reporting Errors	22
CPS Reference Period.....	24
Time Trends.....	26
CPS Imputations	26
Conclusion	28
4. Extrapolating to the Full Data.....	29
The Identification Problem.....	29
Imputing the Data.....	32
Stratifying and Adjusting for Covariates	33
Conclusion	36
5. New Estimates of the Uninsured.....	37

Dual Reporting.....	37
Our Approach.....	40
Revised Estimates of Uninsurance, By Demographic Group.....	40
Discussion.....	42
6. New Estimates of Medi-Cal and Welfare Enrollment Rates.....	45
Pooled Results.....	45
Discussion.....	47
7. Conclusion.....	49
Appendix A. Detailed Notes on File Construction and Matching.....	51
A.1. The Raw Data.....	51
A.2. Matching the CPS and the MEDS.....	52
A.3. Verifying Matches.....	52
Appendix B. Regression Specification for Response Errors.....	55
B.1. The Basic Approach.....	55
B.2 Detailed Logistic Regression Results.....	57
Bibliography.....	61

List of Tables

Table 2.1. Medi-Cal Enrollment in California (millions of persons)	6
Table 2.2. Chronology of CPS-ADS Changes and Effect on Health Insurance Coverage	10
Table 2.3. Reporting Rates (CPS/MEDS)	12
Table 3.1. Sample for Matched Analyses.....	14
Table 3.2. Percentage of People from Full Sample in Final (“Matched”) Sample	17
Table 3.3. Congruence in the Matched Sample between MEDS and CPS Data	18
Table 3.4. Mis-Reporting Rates (in %).....	20
Table 3.5. CPS Reference Period	25
Table 3.6. MEDS Data for CPS Imputed Records.....	27
Table 4.1. Adjustment Factors α	35
Table 5.1. Estimates of Dual Coverage and Uninsurance	38
Table 5.2. Health Insurance Coverage Rates: Unadjusted, Adjusted, Discrepancy, Pooled Years	41
Table 5.3. Health Insurance Coverage Rates: Unadjusted, Adjusted, Discrepancy, 2000 Survey Year/1999 Calendar Year	42
Table 6.1. Enrollment Rates: Unadjusted, Adjusted, Discrepancy Medi-Cal, Pooled Years.....	46
Table 6.2. Enrollment Rates: Unadjusted, Adjusted, Discrepancy Welfare, Pooled Years.....	46
Table 6.3. Enrollment Rates: Unadjusted, Adjusted, Discrepancy Medi-Cal, 2000 Survey/1999 Calendar Year	47
Table 6.4. Enrollment Rates: Unadjusted, Adjusted, Discrepancy Welfare, 2000 Survey/1999 Calendar Year	47
Table A.1. Age Differential (CPS Age – MEDS Age) (Percent within Validation Status)	54
Table B.1. Detailed Logistic Regression Results for Medi-Cal	58
Table B.2. Detailed Logistic Regression Results for Welfare	59

List of Figures

Figure S.1 – Reporting Rates (CPS relative to MEDS) by Age and Program	xiii
Figure 2.1 – Medi-Cal Enrollment, in California, by Age, Welfare and Total.....	6
Figure 2.2 – Under-Reporting of Enrollment (CPS/MEDS)	11
Figure 3.2 – CPS Reporting of Welfare Given MEDS Pattern of Receipt.....	23
Figure 5.1 – Dual Coverage Rates and Adjusting Total Health Insurance Coverage (OHI: Other – non-Medi-Cal – Health Insurance)	38

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.

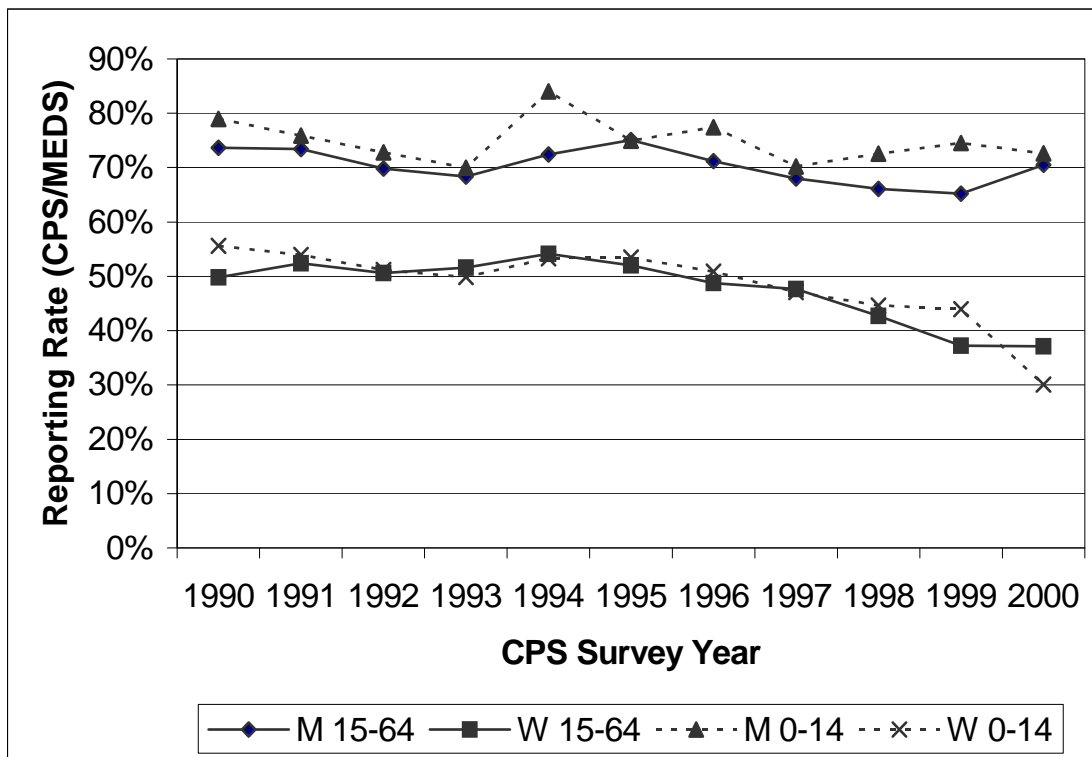


Figure S.1 – Reporting Rates (CPS relative to MEDS) by Age and Program

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.

Acknowledgments

Funding for this analysis was provided by the California HealthCare Foundation (CHCF; through its then separate Medi-Cal Policy Institute—MCPI) and the U.S. Department of Health and Human Services, Administration for Children and Families (DHHS-ACF). The project officers at CHCF/MCPI—Ingrid Aguirre Happoldt, and at DHHS-ACF—Audrey Mirsky-Ashby, Laura Chadwick, and Leonard Sternbach—have been waiting patiently for these results and we appreciate their interest and patience.

This analysis is based on a unique dataset constructed by matching confidential Census Current Population Survey data to confidential administrative data on the Medi-Cal program. Doing so has required the cooperation of several groups. Gene Hiehle and the California Department of Health Services provided the Medi-Cal administrative data and have been supportive throughout this project. B.K. Atrostic and the U.S. Bureau of the Census's Center for Economic Studies provided the Current Population Survey data and handled the matching and deidentification tasks. Senior leadership of the California Census Research Data Center, especially, V. Joseph Hotz at UCLA and Andrew Hildreth at UC Berkeley, have provided crucial support during the negotiations. Center staff, especially Nelson Lim and Becky Acosta, have provided guidance and support with writing the proposal, using the Center, and doing the analyses. The contribution of each of these groups has been necessary in order to gain access to the data.

This research has been presented at RAND, at the California Health Care Foundation, and at the Population Association of America. Comments received in each of these forums have improved the paper. An anonymous RAND interal reviewer also provided very useful comments.

This research is an outgrowth of work begun under the RAND Statewide CalWORKs Evaluation. The constructive comments and guidance provided by CDSS employees during that effort has benefited this effort greatly (though they have not always agreed with our findings). They include Werner Schink, Lois van Beers, Nikki Baumrind, Wilistine Sayas, Aris St. James, and Paul Smilanick.

The data analysis for this project is based on the data preparation work of programmers in RAND's Research Programming Group under the direction of Jan Hanley, who led the effort and did much of the data preparation work herself. Beth Roth, an author on this report, is a member of that group. Other programmers involved in the effort included Christine DeMartini, Laurie McDonald, and Deborah Wesley.

Finally, at RAND this work proceeded within the Labor and Population Program's Center for the Study of Social Welfare Policy. Further information about RAND, the Labor and Population program, and the Center for the Study of Social Welfare Policy can be found at [/www.rand.org](http://www.rand.org), [/www.rand.org/labor](http://www.rand.org/labor), and [/www.rand.org/socialwelfare](http://www.rand.org/socialwelfare). The strong support of RAND, the Labor and Population Program, and its current and former Directors, Arie Kapteyn and Lynn Karoly (respectively), and Assistant Director Rebecca Kilburn, for this effort is gratefully acknowledged. Within RAND, this report has also benefited from the secretarial support of Christopher Dirks and Natasha Kostan. Finally an anonymous reviewer provided technical review improved the final product.

Glossary, List of Symbols, etc.

Symbol	Definition
ADS	CPS March Annual Demographic Survey
AFDC	Aid to Families with Dependent Children
CalWORKs	California Work Opportunities and Responsibility to Kids Act (1997)
CRDC	Census Research Data Center
CCRDC	California Census Research Data Center
CDHS	California Department of Health Services
CDSS	California Department of Social Services
CHCF	California HealthCare Foundation
CHIP	Child Health Insurance Programs (established by statute in 1997, operated by the states)
CMS	Centers for Medicare and Medicaid Services
CPS	Current Population Survey
ESHI	Employer Sponsored Health Insurance
HCFA	Health Care Financing Administration
MEDS	Medi-Cal Eligibility Data System
MCPI	Medi-Cal Policy Institute
NIPA	National Income and Product Accounts
OHI	Other Health Insurance
PIK	Person Identification Key
SIPP	Survey of Income and Program Participation
SSA	Social Security Administration
SSI	Supplemental Security Income
SSN	Social Security Number
TANF	Temporary Assistance to Needy Families

1. Introduction

High-quality survey data are crucial to our understanding of the effects of the nation's social welfare system. If all one wants to know is the number of people participating in a program, then that information can be obtained from administrative data. However, very often, both researchers and policymakers want to know take-up rates (i.e., the fraction of people with certain characteristics enrolled in the program) and the effects of the program on subsequent outcomes (e.g., probability of lacking any health insurance, probability of living in poverty, etc.). For these outcomes, we need richer data that can only be gleaned from surveys; in particular, we need: (1) information on the number and characteristics of nonparticipants; and (2) information on participating families not recorded in administrative data.

Unfortunately, there is considerable evidence that survey data significantly under-report participation in safety-net programs relative to aggregate administrative counts and that the under-reporting has increased over time (e.g, Bavier, 1991). However, most of the evidence to date is based on comparisons between aggregate administrative counts and estimates from survey data. It is our belief that a better understanding of the nature and scope of under-reporting can be obtained by comparing administrative and survey data at the individual level and that is what we seek to do in this report.

This document reports the results of a record-match study of individual-level administrative data for Medi-Cal—the Medicaid program in California, and the Current Population Survey (CPS). With funding from the California HealthCare Foundation (CHCF; through its then separate Medi-Cal Policy Institute—MCPI) and the U.S. Department of Health and Human Services, Administration for Children and Families (DHHS-ACF), and the cooperation of the U.S. Bureau of the Census, the California Department of Health Services (CDHS), and the California Census Research Data Center (CCRDC), we matched administrative data for Medi-Cal from the Medi-Cal Eligibility Data System, (MEDS) to March CPS data for 1990 to 2000. In California, everyone receiving cash assistance (sometimes referred to as welfare)——through Aid to Families with Dependent Children (AFDC), later Temporary Assistance to Needy families (TANF)/California Work Opportunities and Responsibility to Kids (CalWORKs)—is automatically enrolled in Medi-Cal. Since the MEDS administrative data allow us to identify the “type” of Medi-Cal coverage (i.e., why the person is eligible for Medi-Cal), we are able to consider overall Medi-Cal coverage and its two components—welfare and Medi-Cal only (i.e., Medi-Cal, but not welfare)——in our analysis.

Plan of the Report

This report proceeds as follows. The balance of this opening chapter reviews the existing literature on the quality of the CPS data on Medicaid and welfare. The second chapter provides background information on the Medicaid/Medi-Cal program, the MEDS (administrative) data, and the CPS (survey) data. It then characterizes the under-reporting problem, using separate tabulations from each data source. In the third chapter, we turn to the matched data file. For the subset of individuals who provide a valid Social Security Number (SSN), we describe the nature of reporting biases based on a one-to-one match of the survey and administrative data. Unfortunately, not all survey respondents provide an SSN. The fourth chapter provides a technical discussion of our methods for using information from the

matched data to impute welfare and Medi-Cal for the entire California CPS sample. In the fifth chapter, we use the resulting multiply-imputed file to reconsider some of the substantive issues for which the CPS is used. In particular, we explore program take-up by (reported) household income, family structure and other health insurance coverage. The final chapter considers the implications of the results.

Previous Literature on Under-Reporting

The conventional source for information on program take-up is the CPS, the largest annual, national survey. Beginning with the March 1995 CPS, the Census Bureau (Benenfield, 1996a), the Congressional Budget Office (Bilheimer, 1997), General Accounting Office (1997), and the Employee Benefits Research Institute (Fronstin, 1996) each publish annual CPS-based estimates of health insurance coverage and uninsurance. However, the CPS-based estimates of health insurance coverage are much lower and estimates of uninsurance much higher than tabulations from other surveys, such as Survey of Income and Program Participation (SIPP) or the National Survey of America's Families (NSAF) (Bennefield, 1998; Lewis, Ellwood, Czajka, 1998; Fronstin, 2000).¹

In addition and of particular relevance to this study, CPS estimates of Medicaid coverage (Medi-Cal in California) are much lower than corresponding tabulations from administrative data on Medicaid (and Medi-Cal in California), suggesting that survey respondents under-report Medicaid/Medi-Cal coverage. The Urban Institute's TRIM2 model (used by DHHS to simulate program costs) uses administrative data from the Centers for Medicare and Medicaid Services (formerly called the Health Care Financing Administration- CMS/HCFA) to partially correct for such under-reporting. For 1995, this correction for underreporting lowers the fraction of children (0-17) uninsured by 31 percent and the fraction of all non-elderly individuals (0-65) uninsured by 11 percent.²

As part of a discussion of the decline in Medicaid coverage, Ku and Bruen (1999) summarize the national issues and their effect on our understanding of policy.

- 1) "CPS data indicate that about 2.5 million fewer non-elderly people got Medicaid in 1997 than in 1995 (9.3 percent fewer), while administrative data indicate that 1.2 million (3.2 percent) lost Medicaid."
- 2) "CPS data indicate that more children lost coverage than adults from 1995 to 1997, while administrative data indicate [that] the declines were larger for adults."
- 3) "[T]he total number of nonelderly people who had Medicaid at any time in a given year was about 25 to 30 percent lower in the CPS than in administrative counts."
- 4) "[T]here appears to be a growing discrepancy between CPS and administrative data concerning the receipt of benefits like Medicaid, welfare, and food stamps in recent years. . . . Using measures of enrollment during the year, the CPS Medicaid

¹ Other papers focusing on question wording for health insurance items include Rajan et al. (2000), and Nelson and Mills (2001).

² For similar comments about welfare, see <http://www.census.gov/hhes/www/income/assess1.pdf>.

participation estimates were 75 percent of administrative counts in 1995, but fell to 70 percent in 1997.”³

- 5) “Some believe that respondents to the CPS may be reporting their current insurance status, rather than answering the actual question about insurance at any time in the prior year.”

Such reporting biases would cause over-estimates of the number of uninsured Americans and, thus, of the demand for the programs being created by new policy initiatives. Lower than expected enrollment has in fact been a problem (Alpha Center, 2000). While simple reporting bias is unlikely to explain all the lower than expected enrollment, such reporting bias has explicitly been cited by some observers (e.g., Alpha Center, 2000).

The problem of under-reporting is perceived to be so severe that the official U.S. Bureau of the Census report on health Insurance (P60-220, 2002) notes it explicitly and at length in its “Technical Note”:

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.

The problem of under-reporting appears to be particularly severe for welfare. Welfare recipients are categorically eligible for Medicaid. In fact, the CPS imputes Medicaid to anyone who reports receiving welfare. However, welfare reform appears to have worsened reporting of welfare in the CPS, perhaps because of confusion over program names, perhaps because of increased stigma of welfare receipt.

³ Furthermore, concern about the problem has increased. See, for example, Levit et al. (1992, pp. 45-46), reflecting minimal concern about undercounting. “CPS counts of people covered by Medicare and Medicaid programs are reasonably consistent with Health Care Financing Administration (HCFA) program data after allowing for the institutional component missing from CPS.” They compare the 1991 CPS estimate (for 1990) of 24.3 million persons to the HCFA Medicaid program estimate of 25.3 persons. They attribute the difference (only about 4 percent) to “the institutionalized population not included in CPS and difficulties that surveys have capturing Medicaid recipients.” They note that estimates of change over time (in particular 1980 to 1991) are quite similar across the CPS and HCFA data.

See Fronstin (1997), HCFA (1996), and Lewis, Ellwood, and Czjaka (1998) for claims that Medicaid under-reporting has increased.

2. Medi-Cal, the CPS, The MEDS, and Under-Reporting

The core of this project is a data match between administrative data for California's Medicaid program—Medi-Cal (i.e., the MEDS data)—and CPS data. This section begins with a brief description of the Medi-Cal program. It then describes the administrative data (the MEDS) and the survey data (the CPS). Finally, we provide some simple tabulations using the *unmatched* data.

The Medi-Cal Program

Since 1965, Medicaid—a joint federal-state program—has provided health insurance to current welfare recipients and some other qualifying families. During the 1980s and 1990s, coverage was significantly expanded, with particular attention to poor children (often referred to as “the percent programs”) and families that are welfare-eligible, whether they are actually on welfare or not (the 1931(b) program).⁴ California's Medicaid program, Medi-Cal, is a joint effort of the California Department of Health Services (CDHS), which administers the program and handles payments, and the California Department of Social Services (CDSS), which supervises county welfare departments that handle enrollment and re-enrollment.⁵

Figure 2.1 (and Table 2.1) shows Medi-Cal enrollment from the MEDS data (described below) according to the CPS concepts we will use in our main analysis. In particular, we tabulate the total number of individuals enrolled at any time in the calendar year. We distinguish welfare from other Medi-Cal (Medi-Cal only, or simply “MO”). Finally, we consider only the non-elderly, in two groups: those 0-14 as of the following March (who we refer to as “Children”) and those 15-65 as of the following March (who we refer to as “Adults”; we discuss the reason for this child/adult break at 14/15 below).

In the late 1980s, Medi-Cal had just over 4 million enrollees, approximately evenly divided between adults and children, with more welfare than Medi-Cal Only (i.e., during the calendar year at least some Medi-Cal, but never welfare). During the early 1990s, the number of enrollees grew rapidly to over 6 million because of a combination of two factors. First, program eligibility was deliberately expanded. Second, California's deep recession made more people income-eligible, especially through rapid growth in welfare/cash assistance.

From the mid-1990s to the early 2000s, Medi-Cal enrollment has remained relatively stable, near 6 million. This stability is the result of offsetting trends in Medi-Cal sub-programs such as welfare and 1931(b). First, as in the rest of the nation, there has been a sharp drop in welfare/cash assistance since the early 1990s. Second, as was intended (but after a transition period), the 1931(b) program's growth has more than offset the shrinkage in cash assistance. Third, the other components of Medi-Cal such as Supplemental Security Income (SSI), Medically Needy, and “Other” are relatively stable.

⁴ For more discussion of these eligibility changes and their effects, see Gruber (2000).

⁵ For more information on Medi-Cal and its multiple programs see: <http://www.medi-cal.org/> and its fact sheet: <http://www.medi-cal.org/resources/view.cfm?section=Resources&itemID=1397>. For more information on the administration of Medi-Cal, see Klerman and Cox (2003).

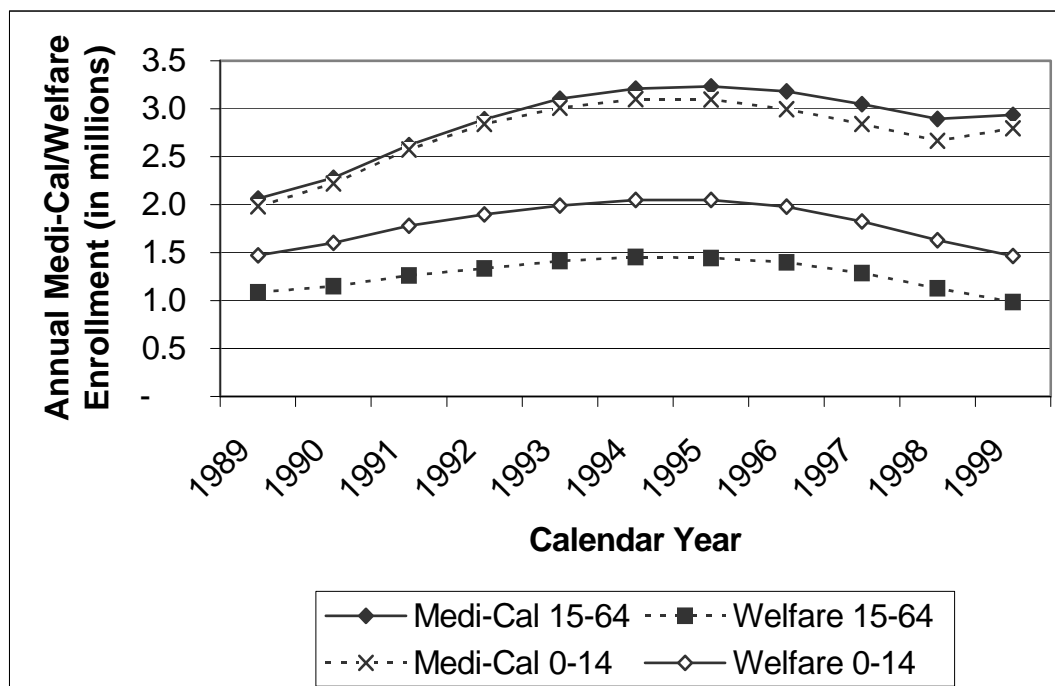


Figure 2.1—Medi-Cal Enrollment, in California, by Age, Welfare and Total

Source: Tabulations from RAND Merged MEDS File

Table 2.1.
Medi-Cal Enrollment in California (millions of persons)

Year	Adults			Children		
	M	W	MO	M	W	MO
1989	2.1	1.1	1.0	2.0	1.5	0.5
1990	2.3	1.2	1.1	2.2	1.6	0.6
1991	2.6	1.3	1.4	2.6	1.8	0.8
1992	2.9	1.3	1.6	2.8	1.9	0.9
1993	3.1	1.4	1.7	3.0	2.0	1.0
1994	3.2	1.5	1.8	3.1	2.1	1.1
1995	3.2	1.4	1.8	3.1	2.0	1.0
1996	3.2	1.4	1.8	3.0	2.0	1.0
1997	3.0	1.3	1.8	2.8	1.8	1.0
1998	2.9	1.1	1.8	2.7	1.6	1.0
1999	2.9	1.0	2.0	2.8	1.5	1.3

Source: RAND tabulations from merged MEDS file.

Note: M-Medi-Cal, W-Welfare, MO-Medi-Cal Only. Tabulated according to CPS concepts: Enrollment is at any time in calendar year (not in each month); Adults are 15-65 in March of the next year; Children are 0-14 in March of the next year.

The Medi-Cal Eligibility Data System (MEDS)

Real-time enrollment information in Medi-Cal is maintained in the MEDS. County welfare departments update this system as individuals are enrolled in or drop out from the program. Providers

check the system to verify whether an individual is covered or not, and which services would be reimbursed by the Medi-Cal program.

Specifically, the MEDS is a person level file using Social Security Number (SSN) as its primary person-level identifier. The data comes to us in overlapping 15-month batches. We link the records using SSN. The quality of the SSN information is expected to be relatively high because the MEDS program has an aggressive program to catch and correct errors in SSN. Moreover, SSN is a primary variable for fraud detection. It is verified initially using the Social Security card and is regularly re-verified. There are some issues with pseudo-SSNs assigned to children prior to their receiving an SSN, but these are also addressed in large part by the state program to identify and fix problem records.

For each month, the file contains an “aid code”. This aid code indicates the program through which the individual receives Medicaid/Medi-Cal coverage (e.g., welfare, SSI, 1931(b), Medically Needy). These codes are of considerable operational relevance. They determine for what medical services a provider will be reimbursed and whether or not a co-pay (referred to as a “share of cost”) must be collected.

We code anyone with a non-zero aid code as receiving Medicaid/Medi-Cal. In the MEDS, the relevant concept is coverage; i.e., is this individual currently holding a Medi-Cal card such that a provider could deliver services with a reasonable expectation of reimbursement. We believe that this corresponds relatively closely to the concept about which the CPS attempts to collect a response. It is possible (in fact, given some ways in which eligibility is conferred and continued, it is likely) that some people are enrolled (or remain enrolled) in Medicaid/Medi-Cal without realizing it. Since we can not directly identify such individuals, we have no choice but to include them. We note that for many (but not all) purposes, they should be included (e.g., if they would be admitted to a hospital for emergency care, the hospital would be reimbursed). Throughout the balance of this document, we use the terms “covered”, “enrolled”, and “participating” interchangeably. In particular, we note that while some authors use “participating” to those who actually receive services, we intend no such distinction.

In California both before and after welfare reform, cash assistance (AFDC/CalWORKs) automatically confers eligibility for (i.e., coverage by) Medicaid/Medi-Cal with the highest level of covered services and no share of cost. As noted above, for operational reasons, this source of coverage is noted in the Aid Code on the MEDS file. It is conventionally used by the state for official purposes (e.g., selecting the Quality Control sample) and by outside analysts to explore welfare related issues. We code welfare corresponding to any AFDC/CalWORKs code providing cash assistance. This includes one-parent and two-parent cases and California’s special refugees and legal immigrant programs (some of which are “state only”). They are referred to as “CalWORKs” in the program names and program documents. In almost all cases, program rules explicitly attempt to treat individuals in all such sub-programs identically. Thus, it seems plausible that CPS respondents in each of these sub-programs would also answer that they were in AFDC/CalWORKs.

Individual-level extracts from the file provide a complete historical record of Medi-Cal eligibility for 1987 forward. Crucially for our purposes, the file includes linking information (name, Social Security Number—SSN), some basic demographics (gender, date of birth, race/ethnicity), and detailed Medi-Cal program information.

In our analysis below, we treat the MEDS records as “truth.” This is a reasonable approximation given their use by providers in determining whether care will be reimbursed. However, the MEDS data are not always absolutely correct. Careful study of the MEDS data suggests some anomalies when

counties had trouble updating the records (e.g., for two months in late 1990, there is a period of a few months when there appear to be no entries onto welfare for Los Angeles County). Card, Hildreth, and Shore-Sheppard note some seam bias (sharp increases in transition rates across versions of the MEDS file), which also suggests some reporting error in the MEDS.⁶

In addition, we note that some people may be enrolled in Medi-Cal but might not be aware of it. In particular, the *Edwards v. Kizer* decision requires California's counties to continue Medi-Cal eligibility for welfare leavers until their eligibility for continued Medi-Cal can be determined. Moreover, California's implementation of the Medicaid 1931(b) program and the provisions of California SB 87 have the effect of keeping many welfare leavers on Medi-Cal even without filing a new application.⁷ It is widely believed that many of these people do not realize they are covered.

The Current Population Survey (CPS)

The CPS is a monthly survey of about 50,000 households conducted by the U.S. Bureau of the Census for the U.S. Department of Labor.⁸ The CPS's primary purpose is to provide official monthly estimates of the unemployment rate, a key business cycle indicator. With its associated sampling weights, it is representative of the American non-institutional population.⁹

Since 1948, in its spring survey the CPS has included additional questions on annual income in the previous year.¹⁰ Today, those additional questions are asked at the end of the March survey (corresponding to the arrival of W-2s and household preparation of federal income tax returns) and are referred to collectively as the Annual Demographic Survey (ADS). Over the years, the set of supplementary questions has grown.

Most important for our purposes, since 1980, the ADS has included detailed questions on health insurance coverage and welfare receipt in the previous calendar year (not as of the date of the March

⁶ See Card, Hildreth, and Shore-Sheppard (2001) for some further discussion of these issues. The seam bias problem should be less severe in the annual reference period of the CPS which we analyze than in the monthly reference period of the SIPP that Card, Hildreth, and Shore-Sheppard analyze. Note also that their biggest matching problems are with children, for whom we do not have SSNs and therefore do not match. Finally, note that below we limit our sample to the validated records which should increase the quality of the SSN data.

⁷ Medicaid Section 1931(b) was a new program created by federal welfare reform (the Personal Responsibility and Work Opportunities Act of 1996) to guarantee Medicaid to any family that would have been eligible for welfare before welfare reform. Section 1931(b) also gave states the option of expanding 1931(b) eligibility to align it with eligibility for cash assistance. California did so with the net effect that welfare leavers with income up to about 165 percent of the poverty line remain indefinitely eligible for Medi-Cal. In practice, implementation of Section 1931(b) in California was delayed until early 1999, but indirect effects (the "Edwards Hold," see Klerman and Cox, 2004) were felt beginning in early 1998.

California SB 87 (chaptered September 30, 2000, effective July 1, 2001) streamlined continued enrollment in Medi-Cal for welfare leavers through adoption of an ex parte process and, in practice, a presumption of continued eligibility for Medi-Cal among welfare leavers. This implementation occurred after the period covered by our data.

⁸ For more on the CPS, see <http://www.bls.census.gov/cps/overmain.htm>.

⁹ The restriction of the CPS universe to the non-institutional population is potentially problematic for analyses of Medi-Cal. While most Medi-Cal enrollees are young, most Medi-Cal expenditures go to the elderly in nursing homes. That group is not in CPS's universe, which is the non-institutional population. The MEDS data do not have a flag for institutional residence. As a partial correction, our analyses below exclude those age 65 and over.

¹⁰ For more on the March Annual Demographic Supplement to the CPS, see:

<http://www.bls.census.gov/cps/ads/adsdes.htm>.

interview).¹¹ These questions began as an attempt to expand the definition of “income” to include employee benefits and non-cash government benefits (Food Stamps, subsidized housing, medical assistance, etc.).

Specifically, in the battery of questions on income receipt, the CPS asks adults about the receipt of public assistance (PAW_YN) by the household in the previous year. Those who respond in the affirmative are then asked the specific form of public assistance (PAW_TYP). We exclude from our welfare variable those who report public assistance, but not AFDC/CalWORKs. This group would include those receiving Supplemental Security Income and those receiving General Assistance (i.e., county level assistance to families without children). This definition corresponds to what is recorded in the MEDS administrative data; i.e., anyone receiving Medi-Cal due to California’s welfare program (called CalWORKs post-welfare reform). See below for a discussion of how we handle imputed and allocated responses (in each analysis).

The definition is, however, problematic for households where some people are on welfare and others are not. In this case, we have followed standard CPS analysis practice and imputed welfare to everyone in the household (more precisely the family or sub-family) when anyone in the household reports welfare. This is likely to lead to false-positives. If someone in the household receives welfare, but this individual does not; we will incorrectly assign them welfare in our CPS analysis file. This is a serious problem with the way the CPS collects its data. It appears unremediable given the available CPS data. We note that from other analyses, it appears that this problem is most salient for child only cases where the parent is an undocumented immigrant (and thus ineligible for welfare), but the child was born in the U.S. (and thus a citizen, eligible for welfare).

Combining the questions on health insurance as an employee benefit with the questions on participation in government health insurance programs yielded a rough measure of total health insurance coverage; its complement provided an estimate of those without health insurance. Until the 2000 interview, the last included in our analysis file, there was no direct question in the CPS about being uninsured. Rather uninsured status is inferred from negative answers to questions about receipt of Medicaid and other types of health insurance (see below Table 2.2 for a summary of changes to the CPS questionnaire).

Specifically, we assign Medicaid/Medi-Cal based on the Census Bureau’s composite Medicaid variable. This variable combines information from the initial CPS question about Medicaid (MCAID—“Was ... covered by medicad) and also a set of follow-up questions among those who otherwise report no other health insurance (OTHYPn). Equivalently, uninsurance is coded based on negative answers to the specific probes for each type of health insurance and to the follow-up questions. See below for a discussion of how we handle imputed and allocated responses (in each analysis).

As this discussion suggests, the individual questions were not originally intended to generate an estimate of the size of the population without health insurance. With issues of uninsurance becoming more salient, in 1988, the Census Bureau refined the questions.¹² Questions about employer-based

¹¹ This discussion draws on Nelson and Mills (2001).

¹² In addition, in 1983, the Census Bureau began a second national survey, the SIPP). The SIPP is a moderate-sized panel survey with more detailed questions on income and program participation (as its name implies). The original vision appears to have been that the SIPP would replace the CPS-ADS for many purposes, including the measurement of health insurance. However, for a variety of reasons (including varying sample size, issues related to

health insurance that previously had only been asked of employed individuals were asked of all individuals 15 or older, regardless of whether they worked. This change should have captured retiree coverage and COBRA benefits (i.e., benefits from a previous employer). In addition, for children, questions were added about health insurance coverage from individuals not residing in the household. This change should have captured coverage provided by non-coresident parents. Finally, the imputation methods for children's coverage were revised and additional questions on Medicaid were added (see Levit et al., 1992; Moyer, 1989; Swartz and Purcell, 1989; and EBRI, 2000).

Additional changes have been made since then. (See also Swartz, 1997; EBRI, 2000.) Table 2.2 presents a detailed chronology. Census analyses suggest that the changes in survey years 1996 and 2000 increased reported health insurance coverage by about 1 percentage point each.

Table 2.2.
Chronology of CPS-ADS Changes and Effect on Health Insurance Coverage

Year	Change	Effect on Health Insurance Coverage
1981	First health insurance questions (employer sponsored and government sponsored) on CPS-ADS	<Not applicable>
1988	Introduction of new CPS processing system	Minimal
1989	Addition of questions on child health insurance coverage (previously coverage of children was imputed based on adult responses)	Moderate
1993	Switch to 1990 Census population controls	Minimal
1994	Switch from paper and pencil to Computer Aided Personal Interviews	Minimal
1996	Questions reordered and modified to improve information on Medicaid	Possibly moderate; see Swartz (1997).
1998	Indian Health Service no longer considered coverage	Minimal
2000	Verification questions added	Moderate (about 1 percentage point)
2001	Switch to 2000 Census population controls	Minimal (less than 1 percentage point)
2001	Addition of questions on state CHIP programs	
2002	Additional sample (78,000 rather than 50,000) to estimate state health insurance coverage rates	Minimal (less than 1 percentage point)

Note: "Year" refers to the survey year. The CPS questions refer to the previous calendar year; i.e. "2001" refers to the survey conducted in March of 2001, collecting information about calendar year 2000. Effect is on the total national coverage rates. Purely because of sampling issues, effects are larger at the state level. Because of substantive issues, effects are often larger for components (e.g., Medicaid). "Minimal" is less than 1 percentage point; "Moderate" is more than one percentage point.

Corresponding to the fact that California has 12 percent of the national population, the annual March ADS to the CPS has about 6,000 California households, about 13,000 individuals (adults and children), and about 2,000 individuals on Medi-Cal.

its panel structure, and slow data release; see Short, 2001), the CPS remains the primary data source for counts of the uninsured. For an analysis similar to this one for the SIPP, see Card et al. (2001).

Table 3.6.
MEDS Data for CPS Imputed Records

CPS Value	MEDS Data		Total
	No	Yes	
Welfare Allocated			
No	94.0%	2.8%	96.8%
Yes	0.76%	2.5%	3.28%
	94.6%	5.4%	100.0%
Medi-Cal Allocated			
No	56.0%	30.3%	86.3%
Yes	6.2%	7.5%	13.7%
	62.2%	37.8%	100.0%
Medi-Cal Logically Imputed			
Yes	51.2%	48.8%	100.0%
Total	51.2%	48.8%	100.0%

Note: Cells are percent within the panel. Verified years only.

Because the CPS does not impute for program participation for many people, the sample sizes for the estimates presented in Table 3.5 are small (about 100 cases are imputed over all of our data). Consequently, these results need to be treated with caution. With that caveat, the results suggest that the hot-deck algorithm is under-estimating program enrollment. For welfare, the difference is small. The MEDS suggests that about 5.4 percent of the allocated cases are enrolled; the Census imputes welfare to only 3.2 percent of them. For Medi-Cal the differences are larger. In MEDS, 37.8 percent of the cases are enrolled in Medi-Cal whereas the CPS imputes only 13.7 percent.

The logical imputations add Medicaid to children whose parents report Medicaid. For older children, we can check this imputation against the MEDS. We find that only 38.8 percent (partially imputed to Medi-Cal) actually have Medi-Cal. The imputation is making things worse.

In the matched sample, the logical imputations are about two percent of all cases and the allocations about one percent of all cases. Relative to perfect imputations, the incorrect logical imputations therefore raise the Medi-Cal enrollment rate by about one percentage point. Relative to perfect imputations, the incorrect allocations lower the Medi-Cal enrollment rate by about a fifth of a percentage point.

For welfare, they represent less than a fifth of a percentage point. The effect on overall enrollments of any imputation errors is therefore trivial. In fact, the imputations of welfare participation appear to be quite good.

These results suggest that people who do not answer the Medicaid questions are substantially more likely to have Medicaid/Medi-Cal than the demographically similar households the CPS hot-deck procedure is using for its imputations. In short, item non-response is not random, but the effect on total estimated enrollment is trivial. In contrast, the Medicaid logical imputations are wrong about half the time, increasing estimated Medi-Cal enrollment by about one percentage point, which is about ten percent of true Medi-Cal enrollment. Some additional attention to the Medicaid logical imputations may be appropriate.

Table B.1.
Detailed Logistic Regression Results for Medi-Cal

	Behavioral				Imputational			
	False Negative		False Positive		False Negative		False Positive	
	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.
INTERCEPT	-0.626	0.129	-4.384	0.156	-3.111	0.129	-0.973	0.119
YEAR_DEV			-0.073	0.017	0.088	0.013		
UNVALID	-0.378	0.079	0.515	0.182				
UNVALID_YR			0.071	0.029				
Y1994			0.169	0.112				
Y1997			0.267	0.124				
Y1998			0.153	0.141				
A1525					0.214	0.103		
A3665			-0.183	0.085	-0.340	0.095		
A4665	-0.445	0.105	0.369	0.103	-0.382	0.137		0.161
M1525					0.872	0.168	-0.648	0.176
M3665					0.801	0.160	-0.626	0.174
M4665							0.229	0.137
MALE	-0.258	0.084	-0.202	0.077	-1.278	0.129	0.530	0.078
HISPANIC	0.707	0.110					0.819	0.138
BLACK	0.367	0.133	0.927	0.122	1.012	0.109	0.431	0.145
HSDO			0.183	0.083				
SOCO	0.504	0.129	-0.168	0.146	-0.532	0.071	0.545	0.108
POV_LT_05	0.366	0.176					0.718	0.118
POV_LT_10	-0.374	0.094	0.579	0.089	0.220	0.083		
POV_LT_15	-0.708	0.112	0.439	0.117	0.169	0.095	-0.429	0.150
POV_LT_20	-0.623	0.147	0.900	0.118	1.169	0.093	-0.937	
KIDSINHH	0.776	0.147	0.531	0.085	0.876	0.079		0.148
SFWKIDS	-0.805	0.131	0.625	0.094	0.352	0.091	-0.529	0.086
OHI	0.819	0.078	-1.167	0.083	-1.094	0.102	0.379	0.088
C_W							-0.414	0.027
A4665_YR	0.027	0.021			-0.045	0.025		
SOCO_YR	0.036	0.026					0.048	0.016
POV_LT_05_YR	-0.095	0.035	0.078	0.026				
POV_LT_20_YR	-0.073	0.024					-0.096	0.024
KIDSINHH_YR	0.065	0.025						
SFWKIDS_YR	-0.025	0.024					-0.036	
OHI_YR					-0.027	0.018		
N		5,961		50,339		50,751		5,549

Table B.2.
Detailed Logistic Regression Results for Welfare

	Behavioral				Imputational			
	False Negative		False Positive		False Negative		False Positive	
	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.
INTERCEPT	0.866	0.294	-7.658	0.318	-5.253	0.164	-1.524	0.238
YEAR_DEV	0.089	0.029	-0.061	0.024				
UNVALID	-0.696	0.220	0.951	0.275	0.018	0.012	-0.467	0.157
UNVALID_YR	-0.104	0.038	0.102	0.046				
A1525	0.530	0.174	-0.953	0.149				
A3665	0.206	0.122						
A4665	-0.230	0.276			-1.127	0.146	0.650	0.352
M1525					0.533	0.106		
M4665					0.782	0.207		
MALE	0.906	0.137			-0.485	0.098		
HISPANIC	0.347	0.091			-0.344	0.096	0.933	0.132
BLACK					0.288	0.153		
HSDO	0.371	0.095	0.364	0.122	0.312	0.084	0.297	0.134
SOCO					-0.527	0.102		
POV_LT_05	0.624	0.214					0.835	0.168
POV_LT_10	-0.046	0.174	0.755	0.147	0.438	0.087		
POV_LT_15			0.503	0.215	0.200	0.115		
POV_LT_20	-0.642	0.166	1.076	0.251	0.973	0.121	-0.687	0.221
KIDSINHH	-0.520	0.232	1.757	0.265	1.709	0.121	0.668	0.178
SFWKIDS	-0.726	0.119	1.922	0.125	0.606	0.091	-0.135	0.063
OHI	1.076	0.131	-1.426	0.164	-0.845	0.088		
C_O					2.017	0.100		
C_O_YR					-0.040	0.018		
HISPANIC_YR					0.053	0.016		
A1525_YR	-0.088	0.032						
A4665_YR	-0.139	0.052						
POV_LT_05_YR	-0.098	0.045						
POV_LT_10_YR	0.084	0.032						
N		3,159		54,141		54,414		1,886

