# THE MEASUREMENT OF MEDICAID COVERAGE IN THE SIPP: EVIDENCE FROM A COMPARISON OF MATCHED RECORDS

David Card
Andrew K. G. Hildreth
Lara D. Shore-Sheppard

Revised December 2002

We thank the editor and two anonymous referees for comments and suggestions on an earlier draft. This project was made possible by the cooperation of the California Department of Health Services (DHS) and the U.S. Census Bureau. Thanks to Gene Hiehle and James Klein of DHS and David Illig of the California Health and Human Services Agency for assistance and guidance in the use of DHS data. We also thank John Abowd, Julia Lane, Ron Prevost, and Lars Vilhuber of the U.S. Census Bureau for their assistance with SIPP data. Barrett Graf provided outstanding research assistance. Funding for part of this project was provided by the California Program on Access to Care (grant number 447615-19900) and by the Joint Center for Poverty Research/U.S. Census Bureau Research Development Grant Program. All opinions are those of the authors and do not reflect the views of the U.S. Census Bureau or the California Department of Health Services.

ABSTRACT

This paper studies the accuracy of reported Medicaid coverage in the Survey of Income and Program Participation (SIPP) using a unique data set formed by matching SIPP survey responses to administrative records from the State of California. Overall, we estimate that the SIPP underestimates Medicaid coverage in the California population by about 10 percent. The probability that a SIPP respondent who is covered by Medicaid in a given month correctly reports their coverage is around 85 percent. The corresponding probability for low-income children is higher – around 90 percent. Under-reporting by those who are actually in the Medicaid system is partially offset by over-reporting of coverage by people who are not. Some of these false positive responses are attributable to errors and missing data in the administrative system, rather than to problems in the SIPP. Taking account of these errors, the estimated false positive rate for the population as a whole is about 1.5 percent, and 4-5 percent for poor children.

David Card
Department of Economics
549 Evans Hall
UC Berkeley
Berkeley, CA 94720-3880
and NBER

Andrew K. G. Hildreth
California Census Research
Data Center
2538 Channing Way
Berkeley, CA 94720-5100

Lara D. Shore-Sheppard
Department of Economics
Williams College
Williamstown, MA 01267
and NBER

One of the most widely debated policy concerns in the United States is the adequacy of health insurance coverage for low-income children and adults. The Medicaid program was established in 1965 to provide health insurance for female-headed families on public assistance and for the aged, blind, and disabled. Over the past two decades the program has gradually expanded to cover low-income families that are not participating in other welfare programs (Gruber, 2002). Despite these expansions, data from the Current Population Survey show that about a quarter of poor children lacked health insurance coverage in the mid-1990s (U.S. Bureau of the Census, 1997). In the wake of recent federal and state-level welfare reforms there have been renewed efforts to maintain and expand Medicaid coverage.[1] Nevertheless, the fraction of the US population with measured health insurance coverage fell slightly over the 1990s.[2]

While analysts agree that expansions in the potential availability of Medicaid have not led to equivalent increases in measured coverage (Shore-Sheppard, 1999; Gruber, 2002), there is less consensus on the reasons for this phenomenon. One simple explanation is that people under-report their true Medicaid status in surveys such as the Current Population Survey (CPS) or the Survey of Income and Program Participation (SIPP) – the two key sources of data on health insurance coverage in the U.S. Indeed, comparisons between administrative data and CPS estimates of the number of Medicaid recipients show a divergence in the 1990s, with much faster

---

[1]Most prominent is the State Children's Health Insurance Program (CHIP). Not all legislative changes have been in the direction of expanding Medicaid coverage: the 1996 federal welfare reform substantially restricted Medicaid eligibility for immigrant children (see U.S. House of Representatives Committee on Ways and Means, 2000, pp. 908-911).

[2]Current Population Survey data show that 86.1 percent of all individuals and 87.3 percent of children under 18 had health insurance coverage in 1992. Similar data for 1999 show 84.5 percent of individuals and 86.1 percent of children covered. (U.S. Census Bureau, 2000).

growth in caseloads than in CPS estimates of the recipient population.[3] Even if under-reporting by people who have coverage is partially offset by "false positive" responses among non-recipients, measurement errors in Medicaid coverage can lead to understatement of the takeup rate for the program, potentially explaining some of the puzzling results in the literature.

In this paper we present new evidence on the accuracy of Medicaid coverage responses in the SIPP. Unlike the March CPS, which asks individuals whether they were covered by Medicaid at any time in the previous year, the SIPP asks questions about coverage on a month-by-month basis. We use a unique data set formed by merging survey information from the 1990-1993 SIPP panels with administrative data on Medicaid eligibility from the State of California's Medi-Cal Eligibility File (MEF). The combined sample contains actual and reported Medicaid eligibility information for 20,000 individuals and 640,000 person-months. We construct estimates of net and gross error rates in reported coverage for the overall population and for various subgroups that can be used by researchers to gauge the potential biases in statistical analyses that use the SIPP data.

The next section of the paper provides a brief overview of the Medicaid program. In Section II we describe the SIPP survey and present a variety of data on measured Medicaid participation patterns in the California sample. Section III describes the eligibility file that provides our administrative data, and summarizes the matching process. We also present information on the characteristics of the matched sample versus the overall California

---

[3]Data from the "HICFA Form 2082" reporting system show that the total number of people covered by Medicaid during the calendar year rose by 28.5 percent between 1992 and 1998, while the average monthly caseload rose by 27 percent (U.S. Department of Health and Human Services Center for Medicaid and Medicare Services, 2002). Data from the March CPS for the same time period show only a 7 percent rise in the number of people covered by Medicaid.

population. Section IV contains our main results, including cross-tabulations of reported Medicaid status in the SIPP survey and the MEF for the overall matched sample, and various subsamples. We also summarize the implications of our findings for studies that use reported Medicaid coverage as either a dependent variable or an explanatory variable. Finally, Section V reviews our main conclusions.

## I. The California Medicaid Program in the Early 1990s

Medicaid is a joint state-federal program that pays for medical services for eligible low-income individuals, including elderly, blind, and disabled recipients of SSI[4]; the "medically needy" (people who have recently incurred large medical expenses); and people in low income families. Historically, the latter group was made up of welfare recipients in the Aid to Families with Dependent Children (AFDC) program. Starting in the mid-1980s, however, a series of federal law changes expanded Medicaid eligibility to families with incomes above the AFDC threshold, and others that did not meet the family composition rules of AFDC. The 1989 Omnibus Budget Reconciliation Act (OBRA) mandated that states offer Medicaid coverage to pregnant women and children up to age 6 with family incomes below 133 percent of the federal poverty threshold. OBRA 1990 further expanded coverage to children born after September 30, 1983 and living in families with incomes below the poverty line. Other legislative changes in the late 1980s and early 1990s allowed states to expand Medicaid coverage beyond these minimum mandates. California, for example, raised the family income limit for pregnant women and infants to 200 percent of the federal poverty line.

---

[4]California operates a state supplemental program known as the State Supplemental Payment (SSP) program that parallels the federal SSI program.

During the 1990s enrollment patterns in the California Medicaid program – known as Medi-Cal – closely tracked national trends. Between 1991 and 1998 the state accounted for a steady 16 percent of average monthly Medicaid enrollment in the U.S. Further, the ratio of per-capita expenditures in California to the nation as a whole remained relatively constant. In light of this stability and the size and diversity of the California population, we believe the state provides an excellent testing ground for evaluating the quality of Medicaid coverage responses.

Table 1 reports the various Medi-Cal eligibility categories in effect in California as of late 1995, along with estimates of the number of people covered under each category.[5] Despite the coverage expansions in the late 1980s and early 1990s, more than three quarters of individuals covered by Medi-Cal in 1995 were adults or children enrolled in AFDC or SSI. The majority of this group – about 60 percent of total Medi-Cal enrollees – consisted of AFDC recipients. Another 10 percent were medically needy adults and children; 5 percent were refugees and undocumented aliens; and 5 percent were medically indigent adults and children. Only about 3 percent of Medi-Cal enrollees in 1995 were women or children who were receiving coverage as a result of the poverty-related expansions.[6]

Given the high fraction of Medi-Cal enrollees whose coverage is linked to welfare participation, it is not surprising that changes in Medi-Cal enrollment are strongly related to changes in the welfare caseload. The recession of the early 1990s led to a rise in California's

---

[5]This table is based on counts of actual Medi-Cal enrollment. People who were potentially eligible for coverage but were not enrolled are not included here.

[6]One reason for the low fraction of the caseload arising from the poverty-related expansions in California is that California has very generous AFDC benefit rates. Consequently, the number of children in families with incomes above the AFDC threshold but below the poverty line is lower than in most states.

welfare rolls and increases in Medi-Cal enrollment. Since 1996 welfare rolls and Medicaid enrollment have both declined in California, with evidence that most of the fall in Medi-Cal has been attributable to the fall in the number of families receiving cash assistance (Broaddus and Guyer, 2000).

## II. Medicaid Coverage Among California Respondents in the 1990-1993 SIPP Panels

In this paper we study the reporting of Medicaid coverage by California respondents in the 1990-1993 SIPP panels. Table 2 provides an overview of the SIPP data. Each panel consists of four rotation groups who are interviewed on a staggered schedule every four months. Individuals in the 1990 and 1991 panels were interviewed 8 times, individuals in the 1992 panel were interviewed 10 times, and those in the 1993 panel were interviewed 9 times. The four panels contain information for a total of 238,938 people covering the period from October 1989 to December 1996. Just over 10 percent of the sample (24,681 individuals) were in California in their first interview. A larger fraction – about 13 percent of the sample or 31,336 people – spent at least one month in California.

Table 3 presents some descriptive statistics for the sample of individuals who were in California at the first SIPP interview, and for various subsets of this population including young children (age 5 or under), all children, people living in poor and "near-poor" families, and people who reported that they were covered by Medicaid in the first survey month. About one quarter of the SIPP California sample are children, and just over 10 percent are elderly. Consistent with national patterns, children are over-represented among the populations of poor and near-poor, and make up close to one-half of Medicaid enrollees. The diversity of the California population is evident in the ethnic composition of the SIPP sample. White non-Hispanics account for under

60 percent of Californians, and make up even smaller fractions of children, people in poverty, and Medicaid recipients. Hispanics (of all racial groups) make up one-quarter of the overall California population, and larger shares of children, people in poverty, and Medicaid recipients.

Like most other longitudinal surveys, the SIPP has sample attrition (see Jabine, King, and Petroni, 1990). About 10 percent of individuals who are in the first SIPP interview leave the sample by the 6th interview, and another 3 percent leave by the 8th interview. Attrition rates are about the same for children as for the overall sample, but are higher for people who were in poor or near-poor families in the first interview, or were enrolled in Medicaid. These selective attrition patterns imply that trends in Medicaid enrollment within a panel are slightly downward-biased relative to trends across panels (see below).

In the early and mid-1990s California had somewhat higher poverty rates than the nation as a whole, and higher welfare recipiency and Medicaid enrollment rates.[7] On average about 15 percent of the SIPP California sample was poor (i.e., had family income below the federal poverty line), with an even higher poverty rate among children. In view of Medicaid eligibility criteria, it is not surprising that over one-half of Medicaid enrollees are poor, and over 85 percent live in families with incomes less than 200% of the federal poverty line.

The bottom rows of Table 3 present Medicaid enrollment data for the SIPP sample. About13 percent of the sample report that they were covered by Medicaid in the first interview month. Enrollment rates in later months are higher, suggesting that coverage rates were rising in the early 1990s. This is confirmed in Figure 1, which plots the fraction of SIPP respondents

---

[7]In the early 1980s California poverty rates were below the national average. By the late 1980s, however, the state's poverty rates consistently exceeded the national average. See Card (2001) for a comparative analysis of labor market and poverty trends in California over the 1980s and 1990s.

living in California who were enrolled in Medicaid by panel and month, along with the average monthly enrollment rate (for all California residents) across all the SIPP panels. There is a tendency for enrollment rates within a panel to rise more slowly than the average enrollment rate. As noted earlier, one explanation for this pattern is that individuals with a higher probability of Medicaid participation have higher attrition rates, so that participation rates are downward biased in later months of the panel. To explore this idea we constructed within-panel changes in Medicaid participation for people who remained in the sample for at least 32 months, and compared these to the changes shown in Figure 1 (which are based on all available observations in each month). As expected, the within-panel changes for the continuing sample are larger, but on average selective attrition can only account for about one-third of the slower within-panel growth in Medicaid participation.

Another striking feature of Figure 1 is that Medicaid coverage rates are higher in the 1993 panel than the two middle panels, and lower in the 1990 panel. We believe that these differences are largely explained by differences in the characteristics of the California sample from panel to panel.[8] In particular, inter-panel comparisons of the fraction of people who live in poverty suggest that the 1990 panel fewer poor families relative to the 1991 and 1992 panels, whereas the 1993 panel has more.[9] Once differences in the distributions of age, ethnicity, and poverty status are taken into consideration, cross-panel differences in Medicaid coverage are

---

[8]The SIPP sample is not designed to be representative of the California population, and some variation will arise in the composition of the sample relative to the underlying population.

[9]We fit a regression for the incidence of poverty to a pooled sample of person-months from the four SIPP panels and included a full set of indicators for the calendar month and dummies for the different panels. Using the 1992 panel as a base, average poverty rates are 3.35% lower in the 1990 panel (standard error 0.19), 0.30% lower in the 1991 panel (standard error 0.15), and 4.00% higher in the 1993 panel (standard error 0.15).

narrowed substantially.[10]

A common feature of longitudinal data collected from retrospective surveys is "seam bias" (see e.g., Jabine, King, and Petroni, 1990; Groves, 1989). SIPP participants are interviewed every four months about their program participation and other activities in the preceding four months. There is a tendency for changes in status to be recorded at the interview "seams" – between the 4th and 5th months for example. Not surprisingly, this is true for changes in reported Medicaid coverage. Figure 2 shows the transition rates into and out of Medicaid coverage in the first 31 months of the combined 1990-1993 SIPP panels. (The sample underlying this figure includes only individuals who were living in California in the current and previous month.) In addition, the figure shows the average fraction of individuals who report Medicaid coverage by SIPP interview month in the pooled sample. Roughly two-thirds of all spell beginnings and endings occur at a seam (versus an expected frequency of 25 percent). Moreover, there is a small but noticeable seam pattern in the rate of Medicaid coverage, with a tendency for higher coverage rates in the month just before the SIPP interview (i.e., interview months 4, 8, 12, ....). These patterns provide prima facie evidence of measurement error in SIPP-reported Medicaid coverage.

Another important indicator of measurement error is the discrepancy between average

---

[10]Specifically, if we fit a regression model for the incidence of Medicaid coverage to a pooled sample of person-months from the four SIPP panels and include indicators for the calendar month and dummies for the different panels, we find that the probability of Medicaid coverage is 1.94% lower (standard error 0.18) in the 1990 panel relative to the 1992 panel, and 2.91% higher (standard error 0.15) in the 1993 panel relative to the 1992 panel. Rates in the 1991 panel are not significantly different from those in the 1992 panel. When we augment the model with controls for ethnicity and interactions of a family poverty indicator with dummies for 5 age ranges, the difference between the 1990 and 1991/1992 panels becomes insignificant, and the difference between the 1993 and 1991/1992 panels narrows to 0.91% (standard error 0.13).

Medicaid participation rates in the SIPP and the average number of Medicaid beneficiaries per capita. Figure 3 shows this comparison. For reference, we also show estimated Medicaid participation rates from the March CPS. The smooth line in the figure is the ratio of monthly Medi-Cal enrollment (for January and July of each year from 1989 to 1996) divided by the Census Bureau estimate of state population in the month – i.e., the "true" fraction of the California population enrolled in Medicaid.[11] The solid line marked with crosses is the average Medicaid enrollment rate from the four SIPP panels. The SIPP-based estimate of Medicaid enrollment tracks the administrative estimate fairly well, but is systematically lower. Taking an average over all months from October 1989 to December 1996 (weighted by the number of people in the combined SIPP California sample in each month) we estimate that the SIPP sample under-estimates Medicaid enrollment by 12 percent.

In contrast to the SIPP survey, which asks about Medicaid enrollment on a monthly basis, the March CPS asks about Medicaid enrollment *at any time* in the previous calendar year. If answered correctly, this rate should be substantially above the average monthly enrollment rate, because many people move in and out of Medicaid during the year. Indeed, using Medicaid administrative data, we estimate that the number of people who were on Medicaid at any time during the year is 25-30 percent higher than the average monthly caseload.[12] However, most analysts have concluded that March CPS respondents report something closer to their current

---

[11]The Census Bureau provides state population estimates for July 1 of each year, and we linearly interpolate to estimate populations as of January 1. The population estimates are based on 1990 Census baselines. Comparisons of these estimates to the 2000 Census population for California reveal a relatively small prediction error (about 1 percent) by the end of the decade.

[12]For example, in the federal fiscal year 1994 (from October 1993 to September 1994) the total number of people who were ever on Medicaid was about 27 percent higher than the average monthly caseload over the year (both nationally and in California).

Medicaid status, rather than their participation at any time in the previous year (e.g., Bennefield, 1996). Consistent with this interpretation, the March CPS enrollment rates in Figure 3 are about equal to estimated enrollment per capita for January of the same year, although the CPS rates rise more slowly over the 1989-96 period. Because of the difficulty of interpreting the March CPS coverage responses, it is hard to compare the relative accuracy of SIPP versus CPS. Our interpretation is that both surveys contain errors, and that the net under-reporting errors from the SIPP appear to be more stable over time.

## III. Matching SIPP and Administrative Eligibility Data

### a. Potential Matching of Medi-Cal Data to SIPP

To move beyond the simple comparisons in Figure 3 it is necessary to match survey responses on Medicaid coverage with administrative data on actual coverage. We compare 1990-1993 SIPP data for California residents with data from the state's Medi-Cal Eligibility Data system. Records were matched using Social Security Numbers (SSNs). Since not all individuals in the SIPP have a valid SSN (or allow the Census Bureau to use their SSN for research purposes), it is important to understand the characteristics of the subsample of SIPP respondents who are eligible for matching.

In the first SIPP interview household respondents are asked to provide names and SSNs for all people in the household. Respondents can provide SSNs, or they can refuse to allow their SSNs to be used, or they respond that they don't have an SSN or don't know it. In subsequent waves the interviewers try to obtain SSNs for individuals who have not yet provided one. Information for respondents who have not explicitly refused the use of their SSN is forwarded to the Social Security Administration for SSN validation. An attempt is made to assign an SSN

(using name, sex, and address information) to respondents whose numbers were not reported.

For purposes of this project a list of validated SSNs for all individuals in the 1990-1993 SIPP panels was searched for matches with SSNs in the administrative file described below. Among the 31,296 individuals in the four SIPP panels who lived in California for at least one month, 76.1 percent had a valid SSN. The fraction is higher (82.2 percent) for the 24,681 people who were living in California at the first SIPP interview. The difference is partially explained by two factors. First, most of the people who are not in California at the start of the panel (93%) join a household that is already in the panel. The "joiners" include new-born infants and young children who are less likely to have a valid SSN.[13] Second, the joiners tend to be in the SIPP for a relatively short time (the median number of months is 10). Thus, interviewers have less chance of obtaining a valid SSN.

Table 4 compares the characteristics of SIPP respondents with and without a valid SSN who were living in California at the first interview. Column 1 reports the characteristics of the overall sample, while columns 2 and 3 report the characteristics of the subsamples with and without a valid SSN, respectively. About a third of the people without an SSN refused to grant permission for the Census Bureau to use their SSN, while the remaining two thirds consists of people who either do not have an SSN, or report an SSN that cannot be validated against their name and address information.

As shown in Appendix Table 1, a key factor determining whether a valid SSN is available is age: only 69 percent of children under the age of 6 (at the first interview) have a valid SSN, compared to 79 percent of 6-24 year olds, and over 85 percent of older adults.

---

[13]About 11 percent of the joiners were born during the panel. Another 13 percent were between the ages of 1 and 16 when they joined the panel.

Because of these differences, the valid-SSN subsample under-represents children. Ethnicity is also a factor: about 86 percent of white non-Hispanics have a valid SSN, compared to 77 percent of black non-Hispanics, 82 percent of Asian non-Hispanics, and 76 percent of Hispanics. Thus, the valid-SSN subsample slightly over-represents white non-Hispanics relative to other groups. Individuals with valid SSNs are also less likely to be poor or near-poor. Sample attrition rates of the subsamples with and without valid SSNs are fairly similar.

Most importantly for this study, Medicaid coverage rates are fairly similar in the subsamples with and without valid SSNs. This equality may seem rather surprising, given that the subsample without SSNs includes a higher fraction of children (who have higher Medicaid coverage rates) and a higher fraction of individuals with low incomes (who also have higher coverage rates). It should be noted, however, that the public assistance and Medicaid systems require SSNs for most people who are covered (except undocumented immigrants). Thus, nearly all adults and children who receive Medi-Cal coverage should have an SSN, or should be in the process of obtaining one. On balance, this requirement offsets the lower rate of SSNs for children and poor adults, leading to a roughly proportional representation of people covered by Medicaid in the valid-SSN subsample.[14] Based on the comparisons in Table 4 we conclude that the group of individuals in the SIPP who can be matched to administrative Medicaid records via their SSNs is not completely representative of the California population, but does include reasonable fractions of children and people from low-income families.

---

[14]Consistent with this argument, the probability of reporting Medicaid in any given month is slightly higher for people with a valid SSN than for those without, once controls for age, ethnicity, and family poverty status are included. For example, in month 32, people with a valid SSN have a 2.4 percent lower probability of coverage without controlling for other factors, but a 2.2 percentage point higher probability with controls.

b. The Medi-Cal Eligibility File

Medicaid enrollment in California is established at county social welfare offices through on-line access to a state-wide database maintained by the state's Health and Welfare Data Center.[15] This file has a record for each individual who is currently enrolled in Medi-Cal, or was enrolled at any time over the previous 15 months. Around the 24th of each month a "snapshot" is taken of the eligibility data base: these snapshots are known as the Medi-Cal Eligibility Files (MEFs). (Here, eligibility refers to the fact that people in the system are eligible to have their medical bills paid by Medi-Cal). Each MEF snapshot includes individual characteristics (sex, date of birth, ethnicity, address, Social Security Number) and two key pieces of information regarding Medi-Cal eligibility: an "eligibility code" summarizing eligibility status, and an "aid code" identifying the program that provides coverage.[16] We use these codes to determine Medicaid coverage status for each person in each month.

For this project the California Department of Health Services granted access to a series of 17 MEFs drawn every six months from July 1989 to July 1997. Each file contains data for the current and previous 12 months. These files provide data for the 109 month period from June 1988 (the earliest date covered by the July 1989 MEF) to July 1997. The files were shipped directly to the U.S. Census Bureau, where all records for individuals with Social Security

---

[15]"Eligibility" as used by the state denotes that an individual is enrolled in the program and may receive services paid for by Medi-Cal.

[16]Medi-Cal, like other state Medicaid systems, offers different types of coverage. Some individuals' expenses are fully covered whereas others have to share costs or spend a certain amount before they are covered. This introduces some ambiguity in the interpretation of Medicaid coverage: individuals who are ineligible until they reach a certain level of expenses could be considered "covered by health insurance" but would be classified as "ineligible" for Medicaid in the MEF.

Numbers that matched those of people in the 1990-1993 SIPP panels were retained, and coded with an identifier that could be matched to the public use versions of the SIPP data. The MEF records for this subset of matched people were then shipped to the California Census Research Data Center, where we matched them to the SIPP files.

An important feature of the MEF records is the overlap in information provided for each person. For example, eligibility information for December 1991 is contained in the January 1992 and July 1992 MEF's, while data for January 1992 is contained in the January 1992 MEF, the July 1993 MEF, and the January 1993 MEF. For a variety of reasons, the information for a given month is not necessarily consistent across MEFs. One important source of inconsistency is that Medicaid eligibility can be established 'after the fact' – this is particularly likely to affect eligibility under medically needy and medically indigent programs (Klein, 1999). A simple way of combining data across MEF files is to adopt the rule that the latest information is "best": thus, eligibility in any given month is assigned based on the last MEF that covers that month. After close examination of the monthly eligibility patterns in overlapping MEF files we decided on a variant of this rule. Specifically, for any calendar month we use the eligibility data in the most recent MEF, with the exception that we did not use the information provided for the 12$^{th}$ previous month.[17]

Even with this rule, the administrative coverage data exhibit a "seam bias" pattern, suggesting that there is some remaining measurement error in assigned coverage. In particular, using MEF records for individuals who can be matched to the SIPP, we estimate that roughly 40

[17]We found many cases where it seemed that eligibility information for the 12$^{th}$ previous month had been over-written with information from the current month. Use of data for the 12$^{th}$ previous month led to a relatively high number of 1-month "gaps" in spells coverage or non-coverage.

percent of spell transitions in any six month period occur at a seam date (months 6, 12, 18,... of the 109 month sample period) versus an expected frequency of 16.7 percent if transitions occurred equally across months. At least some of this seam bias is probably due to our matching procedures. Rather than linking all the MEF records for a single individual prior to attempting a match to the SIPP files, each MEF record with an SSN was linked individually to the SIPP, and the matched records were shipped to the California Census Data Research Center. Consequently, if the SSN was missing on an initial Medi-Cal application, and updated later, only the later MEF records for that individual would be included in the matched file — the early records would be missing. This would make it appear that the individual had started Medi-Cal at the time of the seam. We are unable to determine the magnitude of this source of seam bias since we only have access to MEF records that were successfully matched to a SIPP record. Obviously, however, the MEF dataset contains some errors – a fact that must be taken under consideration in evaluating the reliability of SIPP-reported coverage.

## IV. Analysis of The Matched File

To analyze the accuracy of Medicaid coverage in the SIPP we compare reported coverage status in the SIPP and MEF file for the same individual in the same calendar month. The analysis is restricted to individuals who report a valid SSN and who were living in California (according to SIPP records) in the month in question. Appendix Table 2 provides a summary of the resulting sample. In brief, there are 23,850 individuals in the 1990-1993 SIPP panels who reported living in California in at least one month, and who provided a valid SSN. On average, each person in this sample has just under 28 months of valid SIPP interview data, and provided just under 27 months of data while living in California, leading to a total of

642,859 person-months of potential Medicaid coverage. The average fraction of months with reported Medicaid coverage is 14.02 percent.

For individuals who match with a MEF record, we derive MEF-based coverage for each month directly from their administrative data. In months when an individual is not in the MEF system (but is still living in California) we assume that the individual is not covered by Medicaid. Similarly, for individuals who report a valid SSN and are living in California but **never** appear in the MEF's during the period from July 1988 to July 1997, we assume that the individual was never covered by Medicaid.

It is important to note that any errors in the matching process will lead us to under-estimate Medicaid coverage in the MEF. For example, if the wrong SSN is assigned to an individual in the SIPP there is relatively little chance of finding a match in the MEFs and the individual will be coded as uncovered. More importantly, if an individual's SSN is mis-coded or missing in the MEF record that is the potential source of data for the current month, then no match will be found and an individual who is actually covered by Medicaid will be assigned an uncovered status. We discuss some evidence on the extent of this problem below.

Table 5 presents the cross-tabulations of MEF and SIPP Medicaid coverage for the overall sample of California residents with valid SSNs, and for various subsamples, including children, individuals in poor or near-poor families, and children in lower-income families. The table contains two sets of entries – the upper entry in each cell is based on unweighted data, while the lower entry (in italics) is estimated using the first year sample weights for each person to weight their person-month observations. Since the SIPP sample is based on a stratified sampling scheme, the weighted estimates are arguably preferable, although the estimates tend to be quite similar.

The first two columns of Table 5 provide some information on the particular subsample – the fraction of overall person-months contributed by the subsample, and the fraction of person-months for the subgroup as a whole attributable to people with a valid SSN. The latter ratio is over 80 percent for the overall sample, but is lower for children and people in poor families, as would be expected given the results in Table 4 and Appendix Table 1. The next four columns show the fractions of person-months in each of four possible categories: covered by Medicaid in both MEF and SIPP; uncovered in both files; covered in MEF but not in SIPP; and covered in SIPP but not in MEF. On average just over 4 percent of the person-month observations appear in the two conflicting categories. Interestingly, for the entire sample and for almost all subgroups, the fraction of people who report coverage in SIPP but not in MEF **exceeds** the fraction covered in MEF but not in the SIPP. The estimated Medicaid coverage rate for individuals with valid SSNs is therefore higher using SIPP-reported coverage than using the MEF records, as shown in the final two columns of the table.

Given the evidence in Figure 3 that overall Medicaid coverage rates are understated in the SIPP, this is a surprising conclusion. Under the assumptions that the MEF data are accurate and that there are no errors in the matching process, the only explanation is that SIPP respondents without valid SSNs (i.e., those who are not included in Table 5) substantially under-report their coverage. Indeed, if the 80 percent of people with valid SSN's *over-report* their Medicaid coverage in the SIPP, and yet the overall rate of Medicaid participation in the SIPP is 10 percent below the true rate (as suggested in Figure 3) the implied under-reporting rate for people with missing or invalid SSN's has to be over 40 percent.[18]

---

[18]Let $M_V$ represent the fraction of people with valid SSNs who are covered in the MEF, let $M_I$ represent the corresponding fraction for people with invalid SSNs, and let $\rho$ denote the fraction

A more plausible explanation is missing or invalid SSNs in the MEF system. The MEF records have a field indicating the status of the SSN, and about 20 percent have missing or invalid SSNs.[19] Of course, not all these records potentially match to people who report a valid SSN to the SIPP. For example, children without an SSN and undocumented immigrants will not appear in our sample of SIPP records with valid SSNs.[20] Nevertheless, we believe that a significant fraction of Medicaid cases generated by people with valid SSNs in the SIPP have erroneous or missing SSNs in the administrative system.

To understand the implications of such errors for drawing inferences about the reliability of the SIPP data in our matched sample, suppose that the true fraction of person months covered by Medicaid is $\pi$, and that a fraction $\alpha$ of records in the MEF system have missing or incorrect SSNs. Assume that if a person is covered by Medicaid in a given month, the probability he or she reports the coverage is $(1-f_n)$, where $f_n$ is the "false negative" reporting rate. Similarly, assume that if a person is **not** covered by Medicaid in a given month, the probability he or she reports the coverage is $f_p$, where $f_p$ is the "false positive" reporting rate. Then the probability of

---

of people who report a valid SSN to the SIPP. The overall fraction covered by Medicaid in the MEF is $\rho M_V + (1-\rho)M_I$. Similarly, let $S_V$ and $S_I$ represent the fractions of people with valid and invalid SSNs who report coverage in the SIPP. Then the ratio of the Medicaid caseload to the caseload estimated in the SIPP is $(\rho M_V + (1-\rho)M_I)/(\rho S_V + (1-\rho S_I) \approx 1.1$. The fractions $S_V$, $S_I$, $\rho$, and $M_V$ are all observable. Plugging in these numbers provides an estimate of the MEF coverage rate for people with invalid SSN's equal to 26 percent. The implied reporting rate of true coverage by people with invalid SSN's is $S_I/M_I = 0.58$ – an under-reporting rate of 42 percent.

[19]Of the 5.9 million MEF records in the June 1991 file, 7 percent are coded as having "no valid input" for the SSN field, 8 percent indicate there was no SSN at the date of entry into the system, 2 percent have an unvalid SSN, and 2.5 percent indicate that the individual is an undocumented alien. We are grateful to Lars Vilhuber for his assistance in processing the SSN validity codes on the MEF records.

[20]According to Table 1, 5.4 percent of the Medi-Cal caseload in the mid-1990s consisted of refugees and undocumented immigrants.

observing a person-month of coverage in both SIPP and MEF is $\pi(1-\alpha)(1-f_n)$, the probability of observing a person-month of coverage in the MEF but not the SIPP is $\pi(1-\alpha)f_n$, and the probability of observing a person-month of coverage in the SIPP but not the MEF is $(1-\pi)f_p+\pi\alpha(1-f_n)$. Moreover, the ratio of the measured Medicaid coverage rate in the SIPP to the true coverage rate is $(1-f_n) + f_p\ (1-\pi)/\pi$, which is bigger than 1 if $f_p > f_n\ \pi/(1-\pi)$.

Regardless of the value of $\alpha$, the false negative reporting rate ( $f_n$) is identified by the fraction of people who are covered in both SIPP and MEF, relative to the fraction who are in MEF. For a given value of $\alpha$, the other two parameters ($\pi$ and $f_p$) are also identified from the observed fractions with various combinations of SIPP and MEF coverage.[21]

The first 3 columns of Table 6 report estimates of $\{\pi, f_n, f_p\}$ for the overall sample of people in the SIPP with valid SSNs and various subgroups, under the assumption that $\alpha=0$ – i.e., that there are no missing or incorrect SSNs in the MEF. The fourth column shows the implied ratio of Medicaid coverage in the SIPP to the true coverage rate. Ignoring errors in the MEF system, the implied false negative rate is about 15 percent, the implied false positive rate is 2.8 percent, and the ratio of the Medicaid coverage measured in the SIPP to the true coverage rate is 103 percent. Looking across subgroups of the population, the false negative rate is fairly stable at 8-15 percent. By comparison, the implied probability of a false positive SIPP response varies substantially across groups, with a rate up to 25 percent or children under the age of 5 in poor families. Across nearly all groups, the estimated net coverage is over-reported in the SIPP by 3-5 percent.

---

[21]Let $R_{11}$ represent the fraction of cases in SIPP and MEF, let $R_{10}$ represent the fraction of cases in MEF and not in SIPP, and let $R_{01}$ represent the fraction of cases in SIPP and not MEF. Then $f_n=R_{10}/(R_{10}+R_{11})$; $\pi=R_{10}/(f_n(1-\alpha))$, and $f_p=(R_{01}-\alpha\pi(1-f_n))/(1-\pi)$.

Inferences about the false positive rate and the net coverage rate change when allowance is made for the possibility of errors or omissions of SSNs in the MEF system. Columns 7-10 present estimates of the same parameters under the assumption that 10 percent of MEF records (for people with valid SSNs in the SIPP) have missing or incorrect SSNs. Allowing for such errors has no effect on the false negative rate, but leads to a substantial reduction in the false positive rate, an increase in the implied true coverage rate $\pi$, and a reduction in the ratio of coverage in the SIPP to true coverage. Allowing for errors in the MEF also tends to stabilize the estimated false positive rate across subgroups. For example, the estimated false positive rate ranges from 1 to 7 percent in column 8, compared to a range of 3-28 percent in column 2. We believe this is indicative of the plausibility of the assumption of error rates around 10 percent in the MEF.

In addition, assuming 10 percent error rates in the MEF leads to the implication that the ratio of SIPP coverage to true coverage for people with valid SSN's is around 90 percent for all groups in Table 6. For the entire population, evidence in Figure 3 suggests that SIPP-based coverage is equal to about 90 percent of the true coverage rate. If the error rates $f_n$ and $f_p$ are independent of whether a valid SSN is reported to SIPP, we would expect the net coverage rate from the SIPP for people with valid SSNs to also equal 90 percent.

Given the limitations of our matched file, it is difficult to find direct evidence on the fraction of incorrect or missing SSNs in the MEF. However, we have examined likely errors among one group: children whose mothers report Medicaid coverage in both the SIPP and the MEF. Since most mothers who are covered by Medicaid are eligible through AFDC, their children should also be covered by Medicaid. We therefore recalculated the entries in Table 5, assuming that all children who are recorded as covered by Medicaid in the SIPP but not in the

MEF and whose mothers reported Medicaid coverage in both data sets are actually enrolled in Medicaid. Unfortunately, this procedure only affects the subset of children who are living with mothers who provided valid SSNs to SIPP. Nevertheless, use of this assumption raises the coverage rate for children in the MEF ($\pi$) to a level slightly above the SIPP coverage rate (from 25.4 percent to 26.2 percent, ignoring weights). We believe this provides further confirmation of the likely importance of invalid or missing SSNs in the MEF file, particularly for children.

*Implications of Estimated Misreporting Rates*

Measurement errors in a dichotomous outcome like Medicaid coverage will affect the consistency of conventional statistical estimators when the outcome is used as either an explanatory variable or dependent variable in the analysis. In the case where Medicaid status is used as an explanatory variable, it is conventional to summarize the impact of the measurement errors by the reliability statistic, $\lambda$ (see e.g. Angrist and Krueger, 1999). This is the regression coefficient of true coverage status on observed SIPP coverage status, and is equal to

P(True Coverage="yes" | SIPP="yes") – P(True Coverage="yes" | SIPP="no") .

In terms of our notation, the reliability is

$$\lambda = \pi(1-f_n)/[\pi(1-f_n)+(1-\pi)f_p] - \pi f_n/[\pi f_n+(1-\pi)(1-f_p)] .$$

The reliability index measures the degree of attenuation bias that would arise if observed SIPP coverage status were used as an explanatory variable in a regression model in place of "true" MEF coverage. If other covariates X are included in the regression, and it is assumed that the misreporting rates are constant across the population, then the attenuation bias is

$$\lambda_x = \{ \lambda - R^2/(1-f_n-f_p) \} / \{ 1 - R^2 \} ,$$

where $R^2$ is the R-squared from a linear probability model for observed SIPP coverage status on

the X's (see Card, 1996, equation (4)). The addition of X's that explain Medicaid coverage will lower the effective reliability of the observed indicator.

If observed Medicaid status is used as a dependent variable in the analysis, and it is assumed that the true probability of coverage is $\pi=F(X\beta)$, where F is a cumulative distribution function (e.g., a logistic or normal) then

$$P(\text{SIPP}=\text{"yes"} \mid X) = f_p + (1 - f_n - f_p)\, F(X\beta),$$

(see Hausman, 2001). In the case of a linear probability specification $F(X\beta)=X\beta$, this equation implies that the $\beta$ coefficients will be attenuated by a factor $\mu = (1 - f_n - f_p)$. More generally, if the X's are dummies indicating mutually exclusive categories, the implied probability differences between categories will be attenuated by approximately $\mu$.[22] For example, if $X\beta$ includes a constant and a dummy indicating Medicaid eligibility status, then the estimated takeup rate (the coefficient on the eligibility dummy) will converge in probability to approximately $\mu$ times the true rate.

Columns 5 and 6 of Table 6 report estimates of $\lambda$ and $\mu$ under the assumption that there are no errors in the MEF-recorded SSNs, while columns 11 and 12 report parallel estimates assuming a 10 percent error rate (i.e., $\alpha=0.1$). The estimated reliability of SIPP-reported Medicaid coverage for the California population ranges from 0.80 when MEF errors are ignored to 0.89 assuming a 10 percent error rate. The estimate of $\mu$ is not as sensitive to assumptions about the error rate in the MEF system, and ranges from 0.82 to 0.84. Looking across subgroups, the estimates of $\lambda$ and $\mu$ are relatively stable when errors in MEF are taken into account, but vary more when these errors are ignored. Given the evidence of errors in the MEF,

[22]In the case of a linear probability or logit specification, the attenuation factor is exact.

and the comparisons of net Medicaid coverage rates under alternative assumptions, we believe that an estimate of $\alpha$=0.1 is plausible. This implies that $\lambda$ and $\mu$ are in the range of 80-90 percent for the overall population and for most subgroups (other than people from relatively high-income families). A value of $\mu$ in this range suggests that measurement errors in coverage are only a small part of the explanation for the relatively low estimated takeup rates that are typically found in studies of the impact of the Medicaid expansions (e.g., Currie and Gruber, 1996; Card and Shore Sheppard, 2002).

We have also calculated the values of the parameters $f_n$, $f_p$, $\pi$, $\lambda$, and $\mu$ for a wider range of values of $\alpha$. An assumed value of $\alpha$>0.16 leads to a negative estimate of the false positive rate for the overall sample: this could be taken as an upper bound on the range of feasible error rates in the MEF. For values of $\alpha$ between 0 and 0.16, the implied value of $\lambda$ ranges between 80 and 95 percent, while the implied value of $\mu$ ranges from 83 to 85 percent. Thus, conclusions about the likely attenuation arising from mis-measured Medicaid coverage data are relatively robust to uncertainty about the error rate in the MEF records.

## V. Conclusions

In this paper we use a unique matched data set for California respondents in the 1990-1993 SIPP panels to assess the validity of Medicaid coverage information in the SIPP. A key finding is that the SIPP provides relatively accurate data on Medicaid coverage for those who are actually receiving it. For the population in the SIPP who have valid Social Security Numbers and can be matched, we estimate that 85 percent of all "person-months" of actual Medicaid coverage are accurately reported. This ratio is even higher for groups with a high likelihood of Medicaid coverage, including children and people in low-income families. Our conclusions on

the accuracy of reported coverage for people who are not actually receiving Medicaid are tempered by the observation that any errors in the matching process between the SIPP and the administrative records will lead to an overstatement of the false positive coverage rate. Making no allowance for such errors, we estimate that 2.5 to 3 percent of people who are not covered by Medicaid report that they are covered in the SIPP. However, making a plausible assumption about the rate of missing and invalid Social Security numbers in the administrative data system, we estimate a much lower rate of false positive responses – 1.5 percent overall, and no higher than 5 percent for poor children. The range of error rates in our study suggest that when reported Medicaid coverage from the SIPP is used as either a dependent or independent variable in a statistical analysis, mis-classification errors cause attenuation biases of no more than 20 percent.

References

Angrist, Joshua and Alan B. Krueger. "Empirical Methods in Labor Economics". In Orley Ashenfelter and David Card, editors, *Handbook of Labor Economics* Volume 3A. Amsterdam: Elsevier, 1999.

Bennefield, Robert L. "A Comparative Analysis of Health Insurance Coverage Estimates: Data from CPS and SIPP." Paper presented to the American Statistical Association Joint Meeting (Chicago, IL), August 6, 1996.

Broaddus, Matt and Jocelyn Guyer. "Losing Ground: Recent Declines in the Medi-Cal Enrollment of Families with Children." Unpublished Working Paper, Center on Budget and Policy Priorities (Washington, D.C), May 2000.

Card, David. "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis." *Econometrica* 64 (July 1996), pp. 957-980.

Card, David. "Welfare Reform and the Labor Market Outcomes of Women." In James Lincoln and Paul Ong, eds., *The State of California Labor*. University of California at Berkeley Institute of Industrial Relations, 2001.

Card, David and Lara Shore Sheppard. "Using Discontinuous Eligibility Rules to Identify the Effects of the Federal Medicaid Expansions on Low Income Children." National Bureau of Economic Research Working Paper 9058 (Cambridge, MA), July 2002.

Currie, Janet and Gruber, Jonathan. "Health Insurance Eligibility, Utilization of Medical Care, and Child Health," *Quarterly Journal of Economics*, Vol 111, 1996, pp.431-466.

Groves, Robert M. *Survey Errors and Survey Costs*. New York: John Wiley, 1989.

Gruber, Jonathan. "Medicaid." National Bureau of Economic Research Working Paper 7829 (Cambridge, MA), August 2000. Revised 2002.

Hausman, Jerry. "Mismeasured Variables in Econometric Analysis" Problems for the Right and Problems from the Left." *Journal of Economic Perspectives* 15 (Fall 2001), pp. 57-67.

Jabine, T., K. King and P. Petroni. "Measurement Error". *SIPP Quality Profile*, Chapter 6. United States Department of Commerce Bureau of the Census, 1990. Available at http://www.sipp.census.gov/sipp/qp/c6.html.

Klein, James. "A Researcher's Guide to Medi-Cal Data." Unpublished Paper, Medical Care Statistics Section, Department of Health Services, October 1999.

Shore-Sheppard, Lara. "Stemming the Tide? The Effect of Expanding Medicaid Eligibility on Health Insurance Coverage", Unpublished Working Paper, University of Pittsburgh Department of Economics, 1999.

State of California Legislative Analyst's Office. "Medi-Cal Program," Sacramento: LAO, January 1995.

United States Department of Commerce Bureau of the Census. "Survey of Income and Program Participation (SIPP) Quality Profile." Available at www.census.gov/sipp/ (undated).

United States Department of Commerce Bureau of the Census. "Heath Insurance Historical Tables." Available at www.census.gov/hhes/hlthins/historic/ (updated December 2000).

United States Department of Health and Human Services, Center for Medicaid and Medicare Services. "HCFA-2082 Report". Available at http://cms.hhs.gov/medicaid/msis/mstats.asp.

United States House of Representatives Committee on Ways and Means. "Green Book." 2000 edition. Available at www.access.gpo.gov/congress/wm001.html.

Table 1: Medi-Cal Eligibility Criteria and Caseload as of 1995

| Eligibility Basis | Number Eligible | Percent of Eligibles |
|---|---|---|
| 1. Categorically Eligible AFDC/SSI Recipients | 4,054,300 | 77.5 |
| Families with dependent children in AFDC | | |
| Aged, blind, and disabled in SSI/SSP | | |
| 2. Women and children in low income families | 202,000 | 3.4 |
| Pregnant women with family income < 185% FPL | | |
| Infants in families with income < 185% FPL | | |
| Children Age 1-6 in families with income < 133% FPL | | |
| Children Born After Sept. 1983 in families with income < 100% FPL | | |
| 3. Undocumented persons and refugees | 282,600 | 5.4 |
| Refugees (aged, blind, disabled, under 19) with family income < 133% of 1991 AFDC level | | |
| Undocumented pregnant women meeting Medi-Cal criteria (pregnancy-related services) | | |
| Undocumented persons meeting Medi-Cal criteria (emergency services only) | | |
| 4. Medically Needy Low income families | 522,500 | 10.0 |
| Families with dependent children and aged, blind and disabled persons with family income < 133% of 1991 AFDC level | | |
| Low income families who have “spent down” to eligibility limits | | |
| 5. Medically Indigent Women and Children | 280,500 | 5.4 |
| Pregnant women and children up to age 21 with family income < 133% of 1991 AFDC level, or who have “spent down” to eligibility limits | | |
| 6. TOTAL | 5,230,800 | 100.0 |

Source: State of California Legislative Analyst’s Office (1995). FPL denotes the federal poverty line for the family unit.

Table 2: Characteristics of 1990-93 SIPP Panels

| | 1990 | 1991 | 1992 | 1993 |
|---|---|---|---|---|
| Number of Interviews | 8 | 8 | 10 | 9 |
| Number of Months of Coverage | 32 | 32 | 40 | 36 |
| Earliest Month of Coverage | 10/89 | 10/90 | 10/91 | 10/92 |
| Latest Month of Coverage | 7/92 | 7/93 | 3/95 | 12/96 |
| Number of People Ever in Panel | 69,432 | 44,373 | 62,412 | 62,721 |
| Number Ever in California | 9,200 | 5,806 | 8,081 | 8,249 |
| [percent of total] | [13.3] | [13.1] | [12.9] | [13.2] |
| Number in California at First Interview | 7,213 | 4,707 | 6,307 | 6,454 |
| [percent of total] | [10.4] | [10.6] | [10.1] | [10.3] |

Note: Based on authors' tabulations of SIPP full panel research files.

Table 3: Characteristics of California Residents in First Interview of 1990-1993 SIPP Panels

| | Status in First Interview Month: | | | | | |
|---|---|---|---|---|---|---|
| | All | Under Age 6 | Under Age 16 | Below Poverty Line | Below 2*Poverty Line | On Medicaid |
| *Percent With Age:* | | | | | | |
| Under 6 | 11.0 | 100.0 | 42.7 | 20.1 | 15.8 | 22.7 |
| Under 16 | 25.8 | 100.0 | 100.0 | 43.7 | 35.8 | 46.1 |
| 65 and Older | 10.9 | 0.0 | 0.0 | 3.4 | 10.5 | 12.8 |
| *Ethnicity (percent):* | | | | | | |
| White Non-Hispanic | 57.7 | 46.6 | 47.4 | 33.1 | 40.0 | 35.2 |
| Black Non-Hispanic | 6.3 | 7.8 | 7.6 | 9.6 | 7.8 | 15.2 |
| Asian Non-Hispanic | 10.5 | 8.7 | 10.7 | 11.7 | 10.0 | 14.2 |
| Hispanic | 25.5 | 36.9 | 34.3 | 45.6 | 42.2 | 35.4 |
| *Attrition:* | | | | | | |
| In Survey to Month 12 | 97.6 | 98.2 | 98.7 | 95.8 | 96.1 | 97.4 |
| In Survey to Month 24 | 89.3 | 90.0 | 90.5 | 85.2 | 85.8 | 86.4 |
| In Survey to Month 32 | 84.8 | 87.3 | 87.8 | 81.1 | 82.2 | 82.6 |
| *Ratio of Family Income to Poverty Line:* | | | | | | |
| Under 1.0 | 15.1 | 28.2 | 26.3 | 100.0 | 44.4 | 52.8 |
| Under 2.0 | 34.9 | 50.1 | 48.4 | 100.0 | 100.0 | 85.6 |
| Received AFDC During Month | 6.9 | 17.5 | 15.6 | 31.5 | 18.0 | 53.8 |
| *Medicaid Coverage:* | | | | | | |
| On Medicaid Month 1 | 12.7 | 26.3 | 22.8 | 43.5 | 31.3 | 100.0 |
| On Medicaid Month 12 | 13.9 | 28.6 | 24.9 | 47.2 | 33.8 | 86.7 |
| On Medicaid Month 24 | 13.9 | 28.1 | 24.4 | 45.8 | 33.6 | 80.1 |
| On Continuously Months 1-12 | 10.5 | 20.7 | 18.5 | 38.3 | 26.5 | 82.6 |
| On Continuously Months 1-24 | 8.6 | 17.2 | 15.3 | 32.4 | 22.5 | 69.9 |
| Number of Observations | 24,681 | 2,934 | 6,771 | 4,028 | 8,895 | 3,206 |

Source: Authors' tabulations of SIPP microdata. Means are weighted by SIPP weight assigned for first year of panel (e.g., 1990 weights are used for 1990 panel).

Table 4: Characteristics of California Residents in First Interview of 1990-1993 SIPP Panels with and Without Valid SSN

| | Mean Characteristics | | |
|---|---|---|---|
| | All | With SSN | Without SSN |
| *Percent with Age:* | | | |
| Under 6 | 11.1 (0.2) | 9.0 (0.2) | 21.3 (0.6) |
| Under 16 | 25.8 (0.3) | 22.9 (0.3) | 40.8 (0.7) |
| *Ethnicity (percent):* | | | |
| White Non-Hispanic | 57.7 (0.3) | 60.2 (0.3) | 44.6 (0.7) |
| Black Non-Hispanic | 6.3 (0.2) | 5.9 (0.2) | 8.6 (0.4) |
| Asian Non-Hispanic | 10.5 (0.2) | 10.4 (0.2) | 11.0 (0.5) |
| Hispanic | 25.5 (0.3) | 23.5 (0.3) | 35.9 (0.7) |
| *Attrition:* | | | |
| In Survey to Month 12 | 97.6 (0.1) | 98.0 (0.1) | 95.7 (0.3) |
| In Survey to Month 24 | 89.3 (0.2) | 90.2 (0.2) | 84.8 (0.5) |
| In Survey to Month 32 | 84.8 (0.2) | 85.8 (0.2) | 79.6 (0.6) |
| *Ratio of Family Income to Poverty Line:* | | | |
| Under 1.0 | 15.5 (0.2) | 14.3 (0.2) | 21.8 (0.6) |
| Under 2.0 | 34.9 (0.3) | 33.2 (0.3) | 43.6 (0.7) |
| Received AFDC During | 6.9 (0.2) | 6.9 (0.2) | 6.4 (0.4) |
| *Medicaid Coverage:* | | | |
| On Medicaid Month 1 | 12.7 (0.2) | 13.0 (0.2) | 11.6 (0.5) |
| On Medicaid Month 12 | 13.9 (0.2) | 13.7 (0.3) | 14.7 (0.6) |
| On Medicaid Month 24 | 13.9 (0.2) | 13.6 (0.3) | 15.7 (0.6) |
| On Medicaid Month 32 | 13.8 (0.2) | 13.5 (0.3) | 15.9 (0.6) |
| Number of Observations | 24,681 | 20,281 | 4,400 |

Note: Standard errors in parentheses. Based on authors' tabulations of 1990-1993 SIPP microdata. Means are weighted by SIPP weight assigned for first year of panel (e.g., 1990 weights are used for 1990 panel).

Table 5: Medicaid Coverage Rates in SIPP versus MEF for Individuals with Valid SSNs (Unweighted and Weighted)

| GROUP | Percent Of all Person-months | Percent With Valid SSN | MEF/SIPP Medicaid Coverage: MEF=y SIPP=y | MEF=n SIPP=n | MEF=y SIPP=n | MEF=n SIPP=y | Overall Coverage Rate: MEF | SIPP |
|---|---|---|---|---|---|---|---|---|
| All | 100.00 | 81.04 | 11.59 | 83.98 | 2.00 | 2.42 | 13.59 | 14.01 |
| | *100.00* | *84.28* | *11.43* | *84.69* | *1.69* | *2.19* | *13.12* | *13.62* |
| **Children Only:** | | | | | | | | |
| Age 0-5 | 10.80 | 64.24 | 24.75 | 66.29 | 3.92 | 5.04 | 28.67 | 29.79 |
| | *8.59* | *63.22* | *24.73* | *67.84* | *3.46* | *4.17* | *28.15* | *28.90* |
| Age 6-15 | 15.88 | 77.05 | 20.47 | 73.21 | 3.07 | 3.25 | 23.54 | 23.72 |
| | *15.49* | *77.86* | *20.89* | *73.34* | *2.58* | *3.19* | *23.50* | *24.08* |
| Age 0-15 | 26.67 | 71.86 | 22.02 | 70.71 | 3.38 | 3.90 | 25.40 | 25.92 |
| | *24.09* | *74.42* | *22.15* | *71.47* | *2.87* | *3.51* | *23.50* | *25.02* |
| **By Family Income Status:** | | | | | | | | |
| Poor (<100% Poverty) | 15.79 | 74.52 | 43.64 | 44.59 | 4.86 | 6.90 | 48.50 | 50.54 |
| | *14.92* | *78.18* | *44.29* | *44.62* | *4.43* | *6.66* | *48.72* | *50.95* |
| Near Poor (100-200% Poverty) | 20.69 | 78.16 | 19.28 | 73.21 | 3.40 | 4.11 | 22.68 | 23.39 |
| | *20.42* | *81.64* | *19.95* | *73.22* | *2.99* | *3.83* | *22.95* | *23.79* |
| Higher Income (>200% Poverty) | 63.52 | 83.60 | 2.15 | 95.98 | 0.94 | 0.92 | 3.09 | 3.07 |
| | *64.66* | *86.53* | *2.04* | *96.47* | *0.73* | *0.76* | *2.77* | *2.80* |
| **Lower-Income Children:** | | | | | | | | |
| Age 0-5 and <100% Poverty | 2.98 | 63.71 | 63.69 | 22.27 | 5.53 | 8.51 | 69.22 | 72.20 |
| | *2.37* | *66.48* | *66.39* | *22.08* | *4.61* | *9.92* | *71.60* | *73.31* |
| Age 0-15 and <100% Poverty | 6.85 | 69.28 | 61.00 | 26.54 | 5.08 | 7.38 | 66.08 | 68.38 |
| | *6.10* | *71.88* | *63.21* | *25.65* | *4.21* | *6.93* | *67.42* | *70.14* |
| Age 0-5 and <200% Poverty | 5.53 | 62.38 | 45.09 | 41.40 | 5.53 | 7.98 | 50.62 | 53.07 |
| | *4.36* | *65.83* | *46.56* | *42.03* | *4.89* | *6.52* | *51.45* | *53.08* |
| Age 0-15 and <200% Poverty | 13.23 | 69.07 | 42.48 | 46.09 | 5.07 | 6.36 | 47.55 | 48.84 |
| | *11.79* | *71.69* | *43.67* | *46.22* | *4.34* | *5.78* | *48.09* | *49.45* |

Notes: Entries are percentages of person-months for individuals with valid SSN who are living in California in the month. Unweighted percentages are in regular type, weighted percentages are in italics.

Table 6: Implications of Estimated Coverage Patterns

| | Assuming No Errors in MEF SSN's | | | | | | Assuming 10% Error Rate in MEF SSN's | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | False Neg. | False Pos. | True Cov. Rate | SIPP/ True Cov. | Attenuation Factors: λ | μ | False Neg. | False Pos. | True Cov. Rate | SIPP/ True Cov. | Attenuation Factors: λ | μ |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| All | 14.72 | 2.80 | 13.59 | 103.09 | 0.80 | 0.82 | 14.72 | 1.33 | 15.10 | 92.78 | 0.89 | 0.84 |
| **Children Only:** | | | | | | | | | | | | |
| Age 0-5 | 13.67 | 7.07 | 28.67 | 103.91 | 0.77 | 0.79 | 13.67 | 3.36 | 31.86 | 93.52 | 0.86 | 0.83 |
| Age 6-15 | 13.04 | 4.25 | 23.54 | 100.76 | 0.82 | 0.83 | 13.04 | 1.32 | 26.16 | 90.69 | 0.91 | 0.86 |
| Age 0-15 | 13.31 | 5.23 | 25.40 | 102.05 | 0.80 | 0.81 | 13.31 | 2.02 | 28.22 | 91.84 | 0.89 | 0.85 |
| **By Family Income Relative to Poverty:** | | | | | | | | | | | | |
| <100% | 10.02 | 13.40 | 48.50 | 104.21 | 0.77 | 0.77 | 10.02 | 4.45 | 53.89 | 93.79 | 0.85 | 0.86 |
| 100-200% | 14.99 | 5.32 | 22.68 | 103.13 | 0.78 | 0.80 | 14.99 | 2.63 | 25.20 | 92.82 | 0.87 | 0.82 |
| >200% | 30.42 | 0.95 | 3.09 | 99.35 | 0.69 | 0.69 | 30.42 | 0.71 | 3.43 | 89.42 | 0.77 | 0.69 |
| **Lower-Income Children:** | | | | | | | | | | | | |
| Age 0-5 <100% | 7.99 | 27.65 | 69.22 | 104.31 | 0.68 | 0.64 | 7.99 | 6.21 | 76.91 | 93.87 | 0.76 | 0.86 |
| Age 0-15 <100% | 7.69 | 21.76 | 66.08 | 103.48 | 0.73 | 0.71 | 7.69 | 2.27 | 73.42 | 93.13 | 0.81 | 0.90 |
| Age 0-5 <200% | 10.92 | 16.16 | 50.62 | 104.84 | 0.73 | 0.73 | 10.92 | 6.79 | 56.24 | 94.36 | 0.81 | 0.82 |
| Age 0-15 <200% | 10.66 | 12.13 | 47.55 | 102.71 | 0.77 | 0.77 | 10.66 | 3.48 | 52.83 | 92.44 | 0.86 | 0.86 |

Notes: Entries are based on unweighted entries in Table 5. See text for formulas. Entries in columns 5 and 11, labeled λ, represent the reliability of SIPP-reported coverage. Entries in columns 6 and 12, labeled μ, represent the attenuation in a linear probability model when reported coverage is the dependent variable.

Appendix Table 1: Probabilities of Having Valid SSN for Individuals Living In California in First Month of 1990-1993 SIPP Panels

| Subgroup | Percent with Valid SSN |
|---|---|
| All | 82.2 |
| *By Age:* | |
| Age 5 or less | 68.5 |
| Age 6-15 | 78.9 |
| Age 16-24 | 78.5 |
| Age 24-64 | 86.7 |
| Age 65 or older | 90.7 |
| *By Ethnicity:* | |
| White Non-Hispanics | 85.7 |
| Black Non-Hispanics | 76.6 |
| Asian Non-Hispanics | 81.6 |
| Hispanics | 76.4 |
| *By Poverty Status and Age:* | |
| In poor family, age 15 or less | 72.9 |
| Not in poor family, age 15 or less | 74.9 |
| In poor family, age 16 or older | 78.7 |
| Not in poor family, age 16 or older | 86.0 |
| *By Reported Medicaid Coverage:* | |
| Covered by Medicaid in Mo.1 | 84.1 |
| Not Covered by Medicaid in Mo. 1 | 81.9 |
| Ever Covered by Medicaid | 79.8 |
| Never Covered by Medicaid | 82.3 |

Notes: All table entries are unweighted counts. Sample includes 24,681 people who were living in California in first month of the 1990-1993 SIPP Panels.

Appendix Table 2: Counts of People and Person Months for Individuals in 1990-1993 SIPP Panels Who Were Ever in California

| | Total Sample | Valid SSN | Missing or Invalid SSN |
|---|---|---|---|
| Number of People | 31,336 | 23,850 | 7,486 |
| Average Number of Months in SIPP with valid Medicaid data | 26.16 | 27.91 | 20.58 |
| Average Number of Months in SIPP with valid Medicaid data *and Living in California* | 25.32 | 26.95 | 20.09 |
| Percent of Valid Months in California | 96.79 | 96.56 | 97.62 |
| **Person Months:** | | | |
| Number Person-Months in California (with valid Medicaid data) | 793,283 | 642,859 | 150,424 |
| Number Person-Months in California and Covered by Medicaid | 112,828 | 90,120 | 22,708 |
| Percent of Person-Months Covered by Medicaid | 14.22 | 14.02 | 15.10 |
| **Addendum:** percent of people ever in California with valid SSN | | | 76.11 |
| percent of person-months in California contributed by people with valid SSN | | | 81.03 |

Notes: all table entries are unweighted counts.

**Figure 1: Medicaid Participation Rates in California, By SIPP Panel**

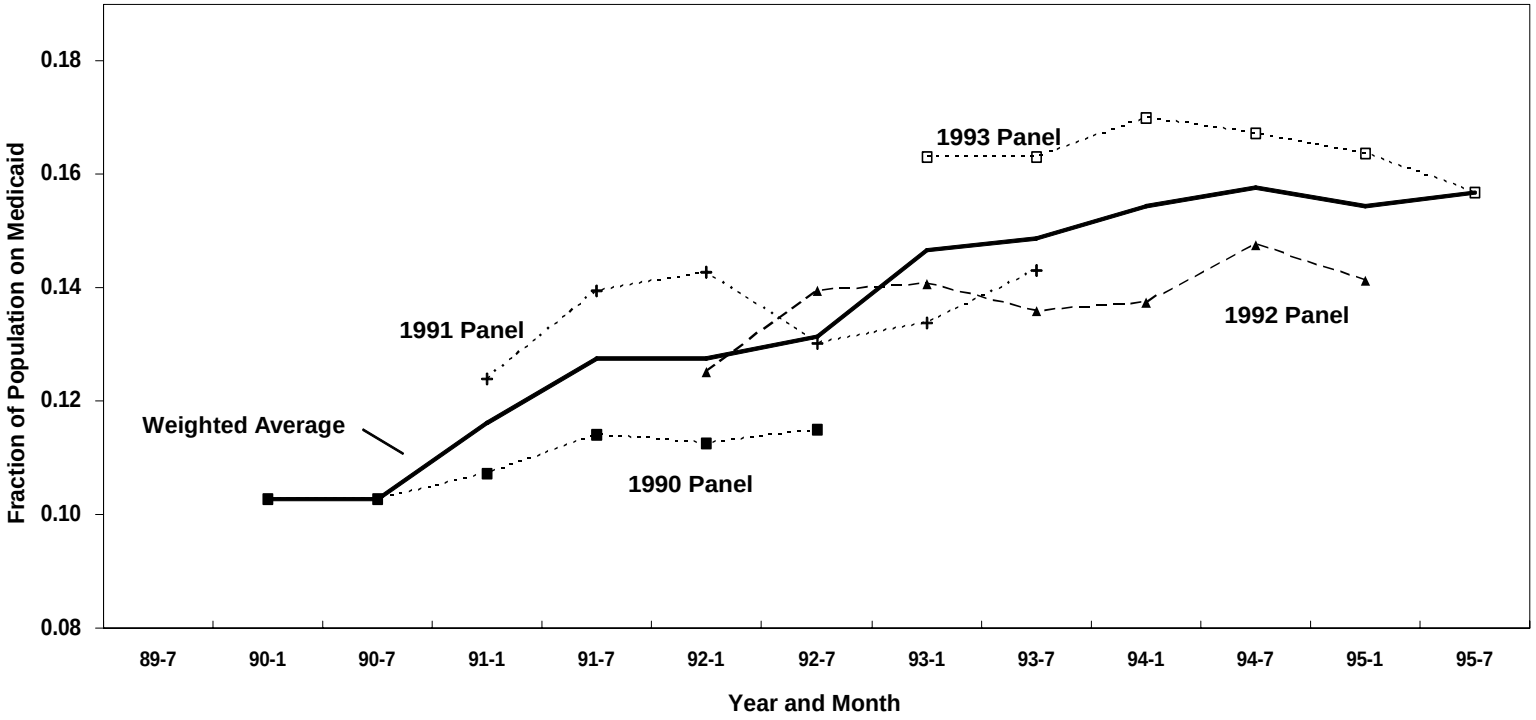

**Figure 2: Medicaid Coverage, Entry and Exit Rates by Survey Month**

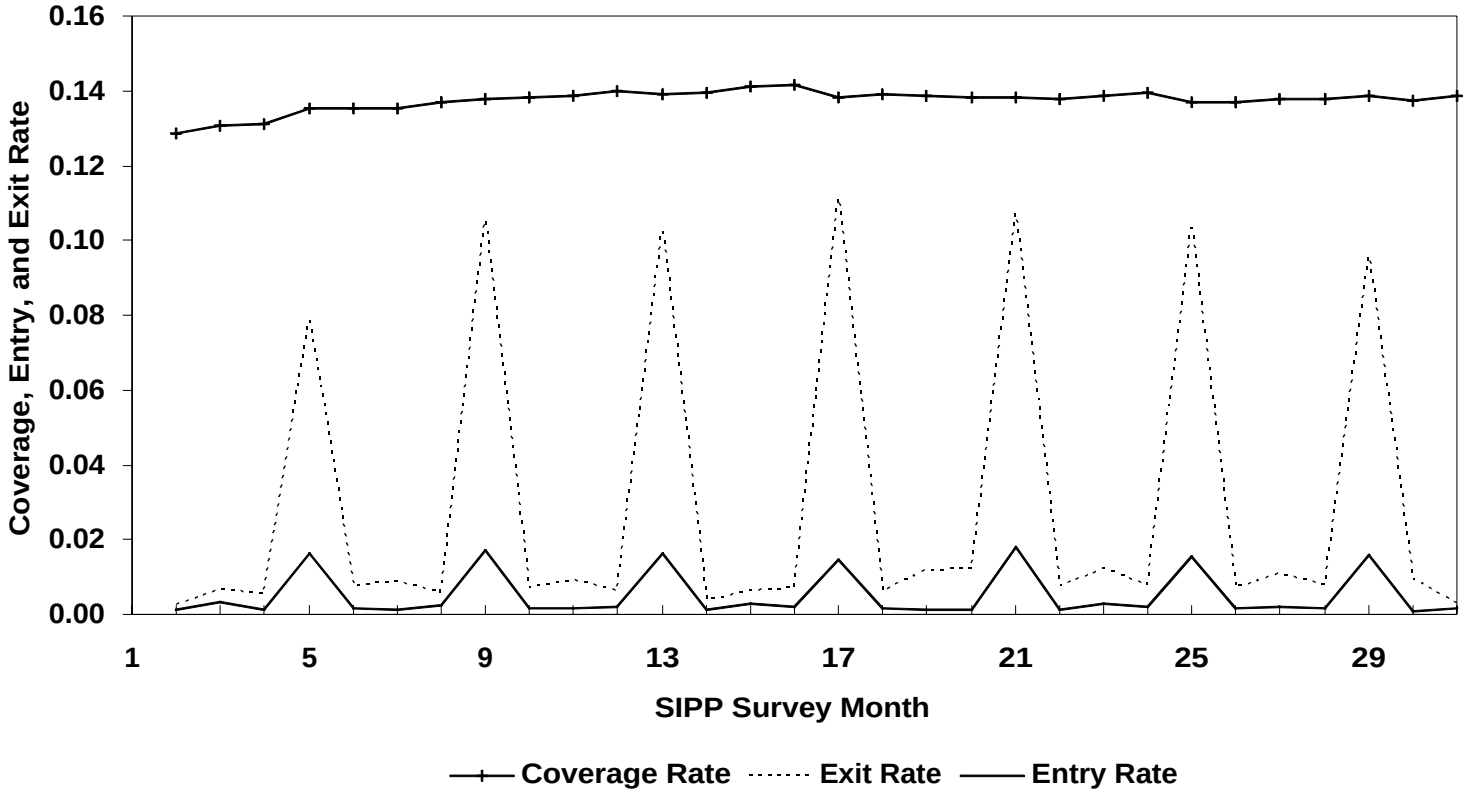

**Figure 3: Medicaid Participation Rates in California, Administration Data, SIPP, and CPS**