

## NBER WORKING PAPER SERIES

THE MEASUREMENT OF MEDICAID COVERAGE IN THE SIPP:  
EVIDENCE FROM CALIFORNIA, 1990-1996

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

Working Paper 8514  
<http://www.nber.org/papers/w8514>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
October 2001

This project was made possible by the cooperation of the California Department of Health Services (DHS) and the U.S. Census Bureau. We are grateful 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. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research, the U.S. Census Bureau or the California Department of Health Services.

© 2001 by David Card, Andrew K. G. Hildreth and Lara D. Shore-Sheppard. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Measurement of Medicaid Coverage in the SIPP:  
Evidence from California, 1990-1996  
David Card, Andrew K. G. Hildreth and Lara D. Shore-Sheppard  
NBER Working Paper No. 8514  
October 2001  
JEL No. I38, C81

### **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. Among SIPP respondents who can be matched to administrative records, we estimate that the probability someone reports Medicaid coverage in a month when they are actually covered is around 85 percent. The corresponding probability for low-income children is even higher – at least 90 percent. These estimates suggest that the SIPP provides reasonably accurate coverage reports for those who are actually in the Medicaid system. On the other hand, our estimate of the false positive rate (the rate of reported coverage for those who are not covered in the administrative records) is relatively high: 2.5 percent for the sample as a whole, and up to 20 percent for poor children. Some of this is due to errors in the recording of Social Security numbers in the administrative system, rather than to problems in the SIPP.

David Card  
Department of Economics  
549 Evans Hall  
UC Berkeley  
Berkeley, CA 94720-3880  
and NBER  
card@econ.berkeley.edu

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 families. Since 1965 Medicaid has been the primary health insurance program for public assistance recipients and other low-income groups. Medicaid was originally designed to provide health insurance for single parent families and the aged, blind, and disabled. Over the past two decades, however, the program has gradually expanded to cover children in lower-income families that are not participating in other welfare programs.<sup>1</sup> Despite these expansions, measured health insurance coverage rates remain far below 100 percent. Data from the Census Bureau's Current Population Survey, for example, show that about 23 percent of people under age 18 who lived in poor households 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.<sup>2</sup> Nevertheless, measured health insurance coverage rates have actually drifted downward over the past decade.<sup>3</sup>

While analysts agree that expansions in the potential availability of Medicaid have not led to equivalent increases in measured coverage (Shore-Sheppard, 1999; Gruber, 2000), there is

---

<sup>1</sup>See Gruber (2000) for a description of these expansions and a review of the literature on their effects.

<sup>2</sup>Most prominent is the State Children's Health Insurance Program (CHIP), implemented in California as the Healthy Families Program. Not all legislative changes have been in the direction of expanding Medicaid coverage: the 1996 federal welfare reforms substantially restricted Medicaid eligibility for immigrant children (see US House of Representatives Committee on Ways and Means Green Book, 2000, pp. 908-911).

<sup>3</sup>In 1992, 86.1 percent of all individuals and 87.3 percent of children under 18 had health insurance coverage. In 1999 the corresponding rates were 84.5 percent of all individuals and 86.1 percent of children. (U.S. Census Bureau, 2000). These data are based on tabulations from the Current Population Survey.

less consensus on the reasons for this phenomenon. One simple explanation is that some people fail to report their true Medicaid status to 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. Available evidence suggests that under-reporting of participation in means-tested government programs is significant. For example, estimates reported in the SIPP Quality Profile (U.S. Bureau of the Census, undated) suggest that SIPP respondents under-report participation in Aid to Families with Dependent Children (AFDC), Food Stamps, and Supplemental Security Income (SSI) by 25 percent. Such a high rate of under-reporting could easily account for the absence of measured health insurance coverage among low-income populations who are eligible for Medicaid.<sup>4</sup>

In this paper we present new evidence on accuracy of reporting of Medicaid coverage in the SIPP among California residents in the early 1990s. Our analysis is based on a unique matched data set composed of survey information from the 1990, 1991, 1992, and 1993 SIPP panels and administrative data on Medicaid eligibility from the State of California's Medi-Cal Eligibility File (MEF). As described below, MEF data were matched to SIPP records for the roughly three quarters of individuals in the SIPP who provided a valid Social Security number (SSN), or granted permission for the Census Bureau to look up their SSN. Using this matched data file it is possible to compare reported versus actual Medicaid eligibility on a month-by-month basis for about 20,000 individuals.

---

<sup>4</sup>For example, Gruber (2000) reports that 32 percent of California children were eligible for Medicaid coverage in 1996, whereas data from the 1993 SIPP show that 29 percent of children (age 16 or younger) were covered by Medicaid in late 1995. If Medicaid coverage is under-reported by 25 percent, however, then the SIPP data actually imply a takeup rate of over 100 percent.

The next section of the paper provides a brief overview of the Medicaid program, with special reference to California. In Section II we describe the SIPP survey and present a variety of data on measured Medicaid participation patterns in the SIPP California sample. Section III describes the eligibility file that is the source of our administrative data, and summarizes the matching process. We also present information on the characteristics of the matched sample versus the overall SIPP California sample. 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, including children. Finally, Section V presents 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<sup>5</sup>; the “medically needy” (people who have recently incurred large medical expenses); and people in low income families. Historically, the latter group was comprised of cash welfare recipients in the 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 all children born after September 30, 1983 and living in families

---

<sup>5</sup>California operates a state supplemental program known as the State Supplemental Payment (SSP) program that parallels the federal SSI program.

with incomes below 100 percent of the poverty line. Other legislative changes in the late 1980s and early 1990s allowed states to expand Medicaid coverage beyond these minimum mandates. Many states responded by raising the income thresholds for coverage: California, for example, raised the family income limit for pregnant women and infants to 200 percent of the federal poverty line. Enrollment patterns in the California Medicaid program – known as Medi-Cal – have closely tracked national trends. During the period from 1990 to 1995 the state accounted for a steady 15 percent of the total Medicaid caseload for 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 relative stability and the size and diversity of the California population, we believe that the state provides an excellent testing ground for evaluating the quality of survey data on Medicaid coverage.

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 and the fraction of the eligible population in each category. 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 who were enrolled in AFDC or SSI. The majority of this group – about 60 percent of total Medi-Cal eligibles – 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 cases in 1995 were women or children who were receiving coverage as a result of the poverty-related expansion programs.<sup>6</sup>

---

<sup>6</sup>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.

Given the high fraction of Medi-Cal enrollees whose eligibility is linked to welfare participation, it is not surprising that changes in Medi-Cal enrollment are strongly related to changes in the welfare caseload. Indeed, as shown in Figure 1, changes in Medi-Cal enrollment over the early 1990s were roughly proportional to changes in the AFDC rolls.<sup>7</sup> The recession of the early 1990s led to a steep rise in California's welfare rolls and similar proportional increases in the Medi-Cal-eligible population. Since the middle of the decade welfare rolls and Medicaid coverage have both declined sharply in California, with evidence that most of the fall in Medi-Cal enrollment has been attributable to the fall in the number of families receiving cash welfare (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, 1991, 1992, and 1993 SIPP panels. Table 2 provides an overview of the four panels. Each SIPP panel consists of four rotation groups who were 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. Overall, the four panels contain information for a total of 238,938 people over 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.

---

<sup>7</sup>Over the 1988-96 period, each additional person in the average monthly AFDC caseload in California was associated with 6 additional Medi-Cal enrollees at some point during the calendar year.

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 6 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 only 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.

An important feature of the SIPP is sample attrition: about 10 percent of individuals who were in the first SIPP interview left the sample by the 6<sup>th</sup> interview, and another 3 percent left by the 8<sup>th</sup> 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. The non-randomness of sample attrition has some implications for our estimates of SIPP response accuracy which we hope to be able to address in future work.

In the early and mid-1990s California had somewhat higher poverty rates than the nation as a whole, and higher welfare reciprocity and Medicaid enrollment rates.<sup>8</sup> On average about 15 percent of the SIPP California sample was poor (i.e., had family income below the federal

---

<sup>8</sup>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.



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. About 13 percent of the sample report that they were covered by Medicaid in the first interview. Consistent with the trends in Figure 1, enrollment is rising within the panels over time. More detail is shown in Figure 2, which plots the fraction of SIPP respondents who were enrolled in Medicaid by panel and month.<sup>9</sup> For reference, we have super-imposed a time line that represents the ratio of average monthly Medi-Cal enrollment in each year to the population of the state – i.e., the “true” fraction of the California population enrolled in Medicaid.<sup>10</sup> The data in Figure 2 suggest that Medicaid enrollment is under-reported in the SIPP. Taking a weighted 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.

A prominent feature of Figure 2 is that average Medicaid enrollment rates vary from panel to panel of the SIPP. For example, the average enrollment rate in January 1993 was 13.3 percent for people in the 1991 panel, 14.0 percent for people in the 1992 panel, and 16.1 percent

---

<sup>9</sup>The rates are weighted coverage rates, using a fixed weight for each individual based on his or her weight for the first calendar year of the appropriate panel. We compare alternative weighting procedures below.

<sup>10</sup>State population is measured as of July 1 from U.S. Census Bureau data. Medi-Cal enrollment is the average monthly number of people who were ever enrolled during the calendar year. Comparisons for 1993-1998 show that average monthly enrollment is very similar to the July enrollment count. One caveat to this calculation is that the state population probably grew faster over the 1990s than was estimated by the Census Bureau.

for people in the 1993 panel.<sup>11</sup> By comparison, the ratio of MEF enrollment to the total California population in January was (approximately) 16.2 percent. We suspect that most of the variation across panels is attributable to variation in the characteristics of people included in the California sample that are correlated with eligibility for Medicaid.<sup>12</sup> In particular, inter-panel comparisons of the fraction of people who live in poverty suggest that the 1990 panel underrepresents poorer families relative to the 1991 and 1992 panels, whereas the 1993 panel overrepresents poorer families.<sup>13</sup> Cross-panel differences in Medicaid coverage rates are narrowed substantially once the comparisons are conditioned on age, ethnicity, and family poverty status.<sup>14</sup>

The strong link between family income and Medicaid coverage is illustrated in Figure 3, which shows the fraction of individuals in different age ranges who report being covered in a given month as a function of their relative family income (the ratio of their family income to the

---

<sup>11</sup>The underlying sample sizes are 4,432, 6,105, and 6,597. Using conventional sampling errors, the rates are significantly different from each other.

<sup>12</sup>The SIPP sample is not designed to be representative of the California population, and some variation will arise across panels in the composition of the sample relative to the underlying population.

<sup>13</sup>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).

<sup>14</sup>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 1992 panels becomes insignificant, and the difference between the 1993 and 1992 panels narrows to 0.91% (standard error 0.13).

appropriate poverty threshold). Apart from the lowest income category, the probability of Medicaid enrollment falls steadily with relative income. As would be expected from the nature of the Medicaid eligibility rules summarized in Table 1, enrollment rates conditional on relative family income are highest for young children, somewhat lower for older children, lower still for adult women, and lowest for adult men.

The rise in Medicaid coverage rates in Figure 3 between the lowest income group (with family incomes under 50% of the poverty line) and the next category (with family incomes between 50% and 75% of the poverty line) illustrates an important feature of the enrollment process. For a variety of institutional and behavioral reasons (including asset limits, spend-down provisions, and the lengthy and complex Medicaid application process) Medicaid enrollment rates vary with both current and “long run” or permanent income. It turns out that families in the SIPP whose current monthly incomes are less than 50% of the poverty line have higher long-run average incomes than those in the next highest income group.<sup>15</sup> Thus, the relatively low Medicaid coverage rate for the lowest income category is explained by the relatively high permanent incomes for people in this category. Apart from this anomaly, average relative incomes over all months of the SIPP are in the same rank order as relative income in any given month, and comparisons across groups based on current monthly income effectively compare Medicaid coverage rates between groups that have similar differences in current *and* permanent relative income.

A well-known feature of longitudinal data collected in retrospective surveys is that

---

<sup>15</sup>The average ratio of monthly family income to the poverty threshold for people in the lowest income category is 1.28, whereas the average ratio for people in the next category is 0.97. Many families that have very low incomes in a month actually report zero or even negative incomes in the month, but much higher incomes in adjacent months.

measured spells of program participation tend to exhibit so-called “seam biases” (see e.g., U.S. Census Bureau, undated, chapter 6; Czajka and Olsen, 2000). SIPP participants are interviewed every four months about activities in the immediately preceding four months. Because individuals have difficulty recalling precise dates in the past (Groves, 1989, chapter 9) and for other unknown reasons there is a tendency for changes in status to be measured at the interview “seams” – between the 4<sup>th</sup> and 5<sup>th</sup> months for example. Not surprisingly, this is true for changes in reported Medicaid coverage. Figure 4 shows the transition rates into and out of Medicaid coverage in each month 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.<sup>16</sup> Roughly two-thirds of all spell starting and ending events 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 most recent months 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.

---

<sup>16</sup>The apparent rise in Medicaid coverage rates in interview months 33-36 is attributable to the influence of observations from the 1993 SIPP panel. Only the 1992 and 1993 panels have data beyond month 32, and only the 1992 panel has data beyond month 36. Medicaid coverage is about 3-4 percentage points higher in the 1993 panel than in other years. The shift in the fraction of data from this panel (from one-quarter in months 1-32 to one-half in months 33-36) would be expected to raise the average coverage rate by 1 percentage point, which is about the increase observed in Figure 4.

### **III. Matching SIPP and Administrative Eligibility Data**

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

Our analysis of reporting errors in Medicaid coverage is based on a matched file composed of 1990-1993 SIPP records and administrative data for individuals who appear in the State of California's Medi-Cal Eligibility Data system at some point between June 1988 and June 1997. 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 SSN to be used, or they respond that they don't have an SSN or don't know it.<sup>17</sup> 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 potentially explained

---

<sup>17</sup>This information is provided by parents or guardians for children.

by two factors. First, most of those who are observed in California at some point but not in the first interview are people who joined a household that is already in the SIPP – only 7 percent were first interviewed in some other state and then followed to California. The “joiners” include new-born infants and young children who are presumably shifting between households.<sup>18</sup> As noted below, children are less likely to have a valid SSN. Second, the joiners tend to be in the SIPP for a relatively short time: the median number of months covered by the SIPP interviews is only 10. Thus, interviewers have less chance of obtaining a valid SSN.

Table 4 compares the characteristics of three samples of SIPP respondents who were living in California at the first SIPP interview. Column 1 reports the characteristics of the overall sample. Column 2 reports the characteristics of the subsample who have a valid SSN, and are therefore potentially eligible to match to the administrative Medicaid records. Finally, column 3 reports the characteristics of the subsample without a valid SSN. Just over one-third of this group consists of people who refuse to allow the use of their SSN, while the remainder consists of people who either do not have an SSN, or report a invalid SSN that cannot be validated using 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 older children and youths (ages 6-24), and over 85 percent of older adults. Because of these differences, the valid-SSN subsample under-represents children relative to the overall sample. Ethnicity is also related to the probability of a valid SSN. About 86 percent of white non-Hispanics have a valid SSN, compared to 77 percent of black

---

<sup>18</sup>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.

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 than those without an SSN. Sample attrition rates, however, are fairly similar for the two groups.

Most importantly for this study, reported 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 those who are covered. Thus, nearly all adults and children who receive Medi-Cal coverage should actually **have** an SSN. On balance, this requirement seems to offset 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. 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.

#### b. The Medi-Cal Eligibility File

Eligibility for Medicaid coverage 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.<sup>19</sup> This file has a record for each individual who is currently eligible for

---

<sup>19</sup>“Eligibility” as used by the state denotes that an individual is enrolled in the program and may receive services paid for by Medi-Cal.

Medi-Cal, or was eligible at any time over the previous 15 months. Around the 24<sup>th</sup> of each month a “snapshot” is taken of the eligibility data base: these snapshots are known as the Medi-Cal Eligibility Files (MEFs). Each monthly MEF includes individual identifying information (sex, date of birth, ethnicity, address, Social Security Number) and two key pieces of information regarding eligibility for Medi-Cal coverage in each month: an “eligibility code” summarizing eligibility status, and an “aid code” giving the type of program the individual is covered by.<sup>20</sup> We use these codes to determine Medicaid eligibility 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 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

---

<sup>20</sup>Medi-Cal, like other state Medicaid systems, offers different types of coverage to different classes of eligible people. 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 eligibility: 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.



1992 and July 1993 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. As noted in Klein (1999), one important source of inconsistency is the fact that Medicaid eligibility can be established 'after the fact' – this is particularly likely to affect eligibility under medically needy and medically indigent programs. One 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.<sup>21</sup> 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<sup>th</sup> previous month.<sup>22</sup> Even with this rule, however, the administrative coverage data exhibit a "seam bias" pattern, suggesting that there is some remaining measurement error in assigned coverage. In particular, a relatively high fraction of transitions into or out of coverage occur at months that represent "seams" between data drawn from different MEFs. Using MEF records for individuals who can be matched to the SIPP, we estimate that roughly 40 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 administrative data seam bias appears to be due to the way the

---

<sup>21</sup>This is the procedure used by State analysts to construct longitudinal coverage histories.

<sup>22</sup>We found many cases where it seemed that eligibility information for the 12<sup>th</sup> previous month had been over-written with information from the current month. Use of data for the 12<sup>th</sup> previous month led to a relatively high number of 1-month "gaps" in spells coverage or non-coverage.

matching between the MEFs and the SIPP public-use identifiers was done. Each record in a MEF that had an SSN was linked individually to the SIPP, and the matched records were shipped to the California Census Data Research Center. Consequently, if an individual received an SSN after beginning Medi-Cal (so that initial MEF records for that individual did not have an SSN but later records did), then later MEF records for that individual would be included on the matched file, but early records would be missing. Thus it would appear in the matched data as if that individual had started Medi-Cal at the time of a seam.<sup>23</sup> This is most likely to have occurred in children, as they are less likely to have an SSN but would be required to obtain one as they participated in Medi-Cal. We are unable to determine the magnitude of this source of seam bias as we have access only to those MEF records that were successfully matched to a SIPP record. However it is likely that the MEF data probably contain some errors – a fact that must be taken into 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 rates 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 brief summary of the resulting sample. Overall, 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

---

<sup>23</sup>Similarly, if the SSN was missing or incorrectly coded in a later MEF but was present and correct earlier, exit from Medi-Cal would appear to have occurred at a seam.

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 recorded as covered by Medicaid 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 appears not to be 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 worth emphasizing that any errors in the matching process will lead us to underestimate 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 an individual who is covered by Medicaid will be incorrectly coded as uncovered.<sup>24</sup> Perhaps more importantly, if an individual's SSN is mis-coded or missing in the MEF, then no match will be found between the SIPP and MEF records, and an individual who is actually covered by Medicaid will be assigned an uncovered status. We discuss some circumstantial 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

---

<sup>24</sup>There is some possibility that an individual with an incorrect SSN will be matched to a MEF record that falsely indicates Medicaid coverage. Since only about 15% of people are covered in any given month, however, a random match is far more likely to lead to a false negative report of coverage than a false positive report of coverage.

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.<sup>25</sup>

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 uncovered in MEF but covered in SIPP. 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 the SIPP but are uncovered in the MEF **exceeds** the fraction who are covered in MEF but fail to report coverage in the SIPP. As a result of this inequality the estimated Medicaid coverage rate for individuals with valid SSNs is higher using SIPP-reported coverage than using the MEF records.

Given the evidence in Figure 2 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 for this pattern is that SIPP respondents without valid SSNs substantially under-report their coverage. As we discuss

---

<sup>25</sup>These weights do not take account of differential probabilities of having a valid SSN, or of differences in sample attrition rates. In future work we plan to implement alternative weighting schemes that adjust for these two phenomena.

below, however, we believe that at least part of the explanation for the lower rate of coverage in the administrative data is incorrect or missing SSNs in the MEF system.

The three right-hand columns of Table 5 present statistics that summarize the accuracy of Medicaid reporting in the SIPP. The first is the probability of reported Medicaid coverage in the SIPP survey for people who are currently covered in the MEF. In the overall sample this fraction is just over 85 percent using the unweighted data, and slightly higher using the weighted data. The second is the probability of reported Medicaid coverage in the SIPP for people who are currently **not** covered in the MEF – this “false positive” response rate is about 2.8 percent for the unweighted sample, and about 2.5 percent using the weighted results. Finally, the last column of the table presents the so-called “reliability” of SIPP-reported Medicaid coverage: this is the regression coefficient of MEF coverage status on observed SIPP coverage status.<sup>26</sup> 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. For the overall sample the reliability of SIPP Medicaid coverage (as a measure of MEF status) is about 80 percent, or slightly higher using weighted data.

Looking at the patterns of SIPP and MEF coverage for children and people in low income families in Table 5, the probability of SIPP-reported coverage for those who are covered in MEF is fairly stable at 85-90 percent. By comparison, the probability of a false positive SIPP response varies substantially across groups, with a rate up to 24-28 percent (depending on the use of weights) for children under the age of 5 in poor families. These results suggest that SIPP-reported coverage is relatively accurate for those who are actually covered by Medicaid.

---

<sup>26</sup>The reliability index is  $P(\text{MEF}=y \mid \text{SIPP}=y) - P(\text{MEF}=y \mid \text{SIPP}=n)$ .

However, the SIPP has a higher rate of errors for people who are not currently covered by Medicaid, but belong to groups with high average probabilities of coverage.

An important caveat to the conclusion that SIPP has a high rate of false positive coverage responses for children is the possibility that MEF coverage status is incorrect. There are two potential sources of errors in MEF coverage status. First, judging by the presence of seam patterns in the coverage transition rates, the MEF records themselves appear to contain some errors. Second, and potentially more important for understanding the apparently high rate of false positive coverage responses in the SIPP, errors in the merging process caused by invalid SSNs in either the SIPP or MEF will lead to cases where an individual is recorded as uncovered in the MEF (because we were unable to find a record for their SSN in the MEF) but is actually covered.

Although we are uncertain about the overall probability of such errors, several factors suggest that the error rate is likely to be higher for children. On the SIPP side, SSNs for children are provided by their parents, and may be less accurately reported than the numbers for adults. On the MEF side, many children enter the MEF system without an SSN and are processed while an application for an SSN is pending. Consequently, early months of participation in Medicaid by such individuals, though correctly reported in the SIPP, will appear to be false positives. In addition, if the SSN is not subsequently updated in the MEF, it will be impossible to find any MEF record for the child, even though the child has a valid SSN and is covered by Medi-Cal.

The MEF contains a field which indicates the validity status of the SSN. Over 80 percent of the 5.9 million SSNs in the MEF in June 1991 are coded as validated. However, 7 percent of MEF records are coded as having “no valid input” for the SSN field, and 8 percent are coded to

indicate that there was no SSN at the date of entry into the system.<sup>27</sup> By comparison with these statistics, virtually all of the MEF records that we can match to individuals in the SIPP are coded as having a validated SSN. The very low match rate for MEF records *without* a validated SSN suggests that the number entered in the SSN field for these records is invalid.

To get a sense of the likely effect of such errors, suppose that 4 percent of all SSNs in the MEF are incorrect, with a 2 percent error rate for adults and a 10 percent error rate for children.<sup>28</sup> Suppose further that 25 percent of the children with incorrect SSNs are actually covered by Medicaid (roughly consistent with the data for children age 0-15 in Table 5) and that 10 percent of all adults with incorrect SSNs are covered by Medicaid. Finally, assume that people with incorrect SSNs report their true coverage status to the SIPP.<sup>29</sup> Under these assumptions, 3.2 percentage points of the person-months recorded as “MEF=no and SIPP=no” should be eliminated from the sample of valid SSN person-months (these are the observations for individuals with incorrect SSNs who are not covered by Medicaid) and another 0.8 percentage points of the person-months recorded as “MEF=no and SIPP=yes” should be eliminated from the sample (these are observations for people with incorrect SSNs who are actually covered by Medicaid). These changes have no effect on the estimated probability of SIPP coverage among those who are covered in the MEF, but they lead to a substantially lower estimate of the

---

<sup>27</sup>Another 2 percent of records have an unvalidated 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.

<sup>28</sup>Anecdotal evidence on the validity of SSNs in other administrative data sets (such as unemployment insurance records) suggests that error rates of 2-6 percent are plausible.

<sup>29</sup>If actual misreporting rates are no higher than 15 percent (as suggested in Table 5) then ignoring misreporting has little effect on the calculations.

probability of SIPP coverage for those without MEF coverage. In particular, the false positive rate for the unweighted data falls to 1.97 percent, versus 2.81 percent in Table 5.<sup>30</sup> Carrying out parallel calculations for children age 0-15, the estimated probability of SIPP-reported coverage for those covered in MEF is again unaffected, but the probability of a false positive SIPP response falls to 2.17 percent (versus 5.22 percent in Table 5).

Given the limitations of the data, it is difficult to determine the true magnitude of matching errors caused by incorrect or missing SSNs in the MEF. However, we are able to examine 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. In Table 6 we show an alternative version of Table 5 in which we have assumed that all children who are recorded as covered by Medicaid in the SIPP but not in the MEF and whose mothers report Medicaid coverage in both data sets are actually enrolled in Medicaid, but erroneously missing from the MEF (either because of missing SSNs or other matching problems). We believe that this is a relatively conservative assumption that gives a lower-bound estimate of the effect of matching errors and missing SSNs. Unfortunately, it only affects some of the children who are potentially missing in the SIPP-MEDS match: those who are living with their mother, whose mothers are themselves covered by Medicaid, and whose mothers also provided valid SSNs to both SIPP and MEDS. Nevertheless, use of this assumption raises the coverage rate for children in the MEF to a level slightly above the SIPP coverage rate. The probability of a false positive response for all children falls by about one-fifth, from 5.22 percent to 4.18 percent. Not surprisingly, the

---

<sup>30</sup>If we double the assumed error rates to 4 percent for adults and 20 percent for children, the implied false positive rate falls to 1.05 percent.



decrease in measured false positive responses is greatest for poor children. Based on these results, we believe that the false positive response rates recorded in Tables 5 and 6 should be viewed as “upper bounds” on the rate of over-reporting of Medicaid coverage in the SIPP.<sup>31</sup>

As noted earlier, the presence of the high rate of false positive coverage responses in Table 5, taken together with the evidence in Figure 2 of overall under-reporting of Medicaid coverage in the SIPP, implies that people **without** valid SSNs must be substantially under-reporting their true coverage to the SIPP. To see this, let  $R_V^*$  represent the fraction of individuals with valid SSNs who are covered by Medicaid according to the MEF, let  $R_I^*$  represent the corresponding fraction among those with invalid SSNs, and let  $\delta$  denote the fraction of people who report a valid SSN to the SIPP. The fraction of the state population covered by Medicaid according to the MEF is  $\delta R_V^* + (1-\delta) R_I^*$ . Similarly, let  $R_V$  and  $R_I$  represent the fraction of individuals with and without valid SSNs, respectively, who report Medicaid coverage in the SIPP. Then the estimated Medicaid coverage from the SIPP for the entire population is  $\delta R_V + (1-\delta) R_I$ . Finally, let  $\lambda$  represent the ratio of the Medicaid caseload measured in the SIPP to the actual number of Medicaid recipients in California. These definitions imply that

$$(1) \quad \delta R_V + (1-\delta)R_I = (\delta R_V^* + (1-\delta) R_I^*).$$

The fractions  $R_V$ ,  $R_I$ , and  $\delta$  are all observable from SIPP data, and  $R_V^*$  is also observable as the MEF-based coverage rate of people in the SIPP with valid SSNs. For the population as a whole the data from Figure 2 suggest that  $\lambda$  is approximately 0.89. Equation (1) can therefore be solved for the “true” rate of Medicaid coverage of people with invalid SSNs in the SIPP ( $R_I^*$ ),

---

<sup>31</sup>As the approach taken in Table 6 will miss child-only AFDC or Medicaid cases, we also tried using a child’s sibling’s report of Medicaid coverage as additional information, however this approach made little difference in the results, probably because there were few children with a sibling who was recorded as receiving Medicaid in both data sets.

and the net under-reporting rate of Medicaid coverage in the SIPP for people with invalid SSNs ( $R_1/R_1^*$ ). Using the unweighted data in the top row of Table 5 and an estimate of  $R_1 = 15$  percent (see Table 4), the implied estimate of  $R_1^*$  is 26.1 percent and the implied estimate of  $R_1/R_1^*$  is 57.6 percent. The very conservative corrections in Table 6 (which affect a limited fraction of children in the sample) reduce the estimate of  $R_1^*$  to 25.1 percent and increase the estimate of  $R_1/R_1^*$  to 59.8 percent.

If an adjustment is made for the possibility of incorrect or missing SSNs for people other than the children whose records were updated in the construction of Table 6, however, the calculations change substantially. For example, assuming that 4 percent of people (10 percent of children and 2 percent of adults) have incorrect SSNs in the MEF, the estimate of  $R_1^*$  falls to 21 percent and the implied estimate of  $R_1/R_1^*$  rises to 71 percent. Doubling the fractions of invalid SSNs to 8 percent (20 percent of children and 4 percent of adults) leads to an implied estimate of  $R_1/R_1^*$  equal to 83 percent. Since  $P(\text{SIPP}=\text{yes} \mid \text{MEF}=\text{yes})$  for people with valid SSNs is fairly stable across groups, we believe it is unlikely that the relative rate of under-reporting of Medicaid coverage in the SIPP for people with invalid SSNs is as low as 55 percent. A more plausible explanation, which is supported by the evidence in Table 6, is that erroneous or missing Social Security numbers in the MEF lead to matching errors, and hence the MEF-based Medicaid coverage rate for people with valid SSNs in the SIPP is understated.

## V. Conclusions

In this paper we use a unique data set, formed by matching California respondents of the 1990-1993 SIPP panels to state administrative records, to assess the validity of reported Medicaid coverage in the SIPP. Our main 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 85-87 percent of all “person-months” of actual Medicaid coverage are accurately reported. This ratio is even higher (at least 90 percent) 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. This rate is substantially higher for children and people from poor families. The apparent false positive error rate for low-income children is reduced by about one-quarter when we use a relatively conservative assumption to update the administrative records of children whose mothers report coverage in both the SIPP and the administrative data. Although such a correction can only be performed for a small subset of the sample, simulations with plausible error rates suggest that the 2.5 to 3 percent rate of false positive coverage responses should be viewed as an upper bound, and that the true rate is probably closer to 2 percent.

## References

- 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. "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.
- 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.
- Czajka, John L. and Cara Olsen. "The Effects of Trigger Events on Changes in Children's Health Insurance Coverage." Unpublished Paper, Mathematica Policy Research, Inc. (Washington, D.C.), April 2000.
- Dorinski, S. M. "Continuing Research on Use of Administrative Data in SIPP Longitudinal Estimation." Working Paper 209, SIPP Data Program, US Bureau of the Census, 1995.
- Ellwood, Marilyn R. and Kimball Lewis. "On and Off Medicaid: Enrollment Patterns for California and Florida in 1995." *Assessing the New Federalism Occasional Paper No. 27*, Urban Institute (Washington D.C.), July 1999.
- Groves, Robert M. *Survey Errors and Survey Costs*. New York: John Wiley, 1989.
- Gruber, Jonathan. "Medicaid", NBER Working Paper 7829, August 2000.
- 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/](http://www.census.gov/sipp/) (undated).
- United States Department of Commerce Bureau of the Census. "Health Insurance Historical Tables." Available at [www.census.gov/hhes/hlthins/historic/](http://www.census.gov/hhes/hlthins/historic/) (updated December 2000).
- United States House of Representatives Committee on Ways and Means. "Green Book." 2000 edition. Available at [www.access.gpo.gov/congress/wm001.html](http://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 Families with dependent children in AFDC Aged, blind, and disabled in SSI/SSP	4,054,300	77.5
2. Women and children in low income families 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	202,000	3.4
3. Undocumented persons and refugees 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)	282,600	5.4
4. Medically Needy Low income families 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	522,500	10.0
5. Medically Indigent Women and Children Pregnant women and children up to age 21 with family income < 133% of 1991 AFDC level, or who have "spent down" to eligibility limits	280,500	5.4
6. TOTAL	5,230,800	100.0

Source: State of California Legislative Analyst's Office (1996). 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 [percent of total]	9,200 [13.3]	5,806 [13.1]	8,081 [12.9]	8,249 [13.2]
Number in California at First Interview [percent of total]	7,213 [10.4]	4,707 [10.6]	6,307 [10.1]	6,454 [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
<i>Percent With Age:</i>						
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
<i>Ethnicity (percent):</i>						
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
<i>Attrition:</i>						
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
<i>Ratio of Family Income to Poverty Line:</i>						
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
<i>Medicaid Coverage:</i>						
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
<i>Percent with Age:</i>			
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)
<i>Ethnicity (percent):</i>			
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)
<i>Attrition:</i>			
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)
<i>Ratio of Family Income to Poverty Line:</i>			
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)
<i>Medicaid Coverage:</i>			
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						Overall Coverage Rate:		Conditional Probability of SIPP Coverage:	
			MEF=Y		MEF=n		MEF	SIPP	MEF=Y	MEF=n	SIPP	Reliability
			SIPP=Y	SIPP=n	SIPP=Y	SIPP=n						
All	100.00	81.04	11.59	83.98	2.00	2.42	13.59	14.01	85.27	2.81	0.80	
	100.00	84.28	11.43	84.69	1.69	2.19	13.12	13.62	87.12	2.52	0.82	
<b>Children Only:</b>												
Age 0-5	10.80	64.24	24.75	66.29	3.92	5.04	28.67	29.79	86.32	7.06	0.77	
	8.59	63.22	24.73	67.84	3.46	4.17	28.15	28.90	87.74	6.80	0.81	
Age 6-15	15.88	77.05	20.47	73.21	3.07	3.25	23.54	23.72	86.95	4.25	0.82	
	15.49	77.86	20.89	73.34	2.58	3.19	23.50	24.08	88.99	4.17	0.83	
Age 0-15	26.67	71.86	22.02	70.71	3.38	3.90	25.40	25.92	86.69	5.22	0.80	
	24.09	74.42	22.15	71.47	2.87	3.51	23.50	25.02	88.53	4.68	0.85	
<b>By Family Income Status:</b>												
Poor	15.79	74.52	43.64	44.59	4.86	6.90	48.50	50.54	89.98	13.41	0.77	
(<100% Poverty)	14.92	78.18	44.29	44.62	4.43	6.66	48.72	50.95	90.90	12.99	0.78	
Near Poor	20.69	78.16	19.28	73.21	3.40	4.11	22.68	23.39	85.01	5.31	0.78	
(100-200% Poverty)	20.42	81.64	19.95	73.22	2.99	3.83	22.95	23.79	86.96	4.98	0.80	
Higher Income	63.52	83.60	2.15	95.98	0.94	0.92	3.09	3.07	69.49	0.95	0.69	
(>200% Poverty)	64.66	86.53	2.04	96.47	0.73	0.76	2.77	2.80	73.79	0.78	0.72	
<b>Lower-Income Children:</b>												
Age 0-5 and	2.98	63.71	63.69	22.27	5.53	8.51	69.22	72.20	92.01	27.64	0.68	
<100% Poverty	2.37	66.48	66.39	22.08	4.61	9.92	71.60	73.31	93.50	23.86	0.73	
Age 0-15 and	6.85	69.28	61.00	26.54	5.08	7.38	66.08	68.38	92.31	21.75	0.73	
<100% Poverty	6.10	71.88	63.21	25.65	4.21	6.93	67.42	70.14	93.76	21.27	0.76	
Age 0-5 and	5.53	62.38	45.09	41.40	5.53	7.98	50.62	53.07	89.07	16.16	0.73	
<200% Poverty	4.36	65.83	46.56	42.03	4.89	6.52	51.45	53.08	90.49	13.43	0.77	
Age 0-15 and	13.23	69.07	42.48	46.09	5.07	6.36	47.55	48.84	89.34	12.13	0.77	
<200% Poverty	11.79	71.69	43.67	46.22	4.34	5.78	48.09	49.45	90.97	11.12	0.80	

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: Medicaid Coverage Rates in SIPP versus MEF for Individuals with Valid SSNs, Medicaid Coverage Assigned if Reported in SIPP and Mother Covered in MEF (Unweighted only)

GROUP	Percent Of all Person-months	Percent With Valid SSN	MEF/SIPP Medicaid Coverage				Overall Coverage Rate:		Conditional Probability of SIPP Coverage:		
			MEF=y SIPP=y	MEF=n SIPP=n	MEF=y SIPP=n	MEF=n SIPP=y	MEF SIPP	Rate: SIPP	MEF=y MEF=n	Reliability SIPP	
All	100.00	81.04	11.82	83.98	2.00	2.20	13.82	14.01	85.51	2.55	0.82
<b>Children Only:</b>											
Age 0-5	10.80	64.24	25.78	66.29	3.92	4.00	29.71	29.79	86.80	5.69	0.81
Age 6-15	15.88	77.05	21.16	73.21	3.07	2.56	24.23	23.72	87.32	3.38	0.85
Age 0-15	26.67	71.86	22.83	70.71	3.38	3.08	26.21	25.92	87.11	4.18	0.84
<b>By Family Income Status:</b>											
Poor (<100% Poverty)	15.79	74.52	44.76	44.59	4.86	5.79	49.62	50.54	90.20	11.49	0.79
Near Poor (100-200% Poverty)	20.69	78.16	19.52	73.21	3.40	3.87	22.91	23.39	85.17	5.03	0.79
Higher Income (>200% Poverty)	63.52	83.60	2.18	95.98	0.94	0.90	3.12	3.07	69.73	0.93	0.70
<b>Lower-Income Children:</b>											
Age 0-5 and <100% Poverty	2.98	63.71	66.05	22.27	5.53	6.14	71.58	72.20	92.28	21.62	0.72
Age 0-15 and <100% Poverty	6.85	69.28	63.34	26.54	5.08	5.04	68.42	68.38	92.57	15.95	0.77
Age 0-5 and <200% Poverty	5.53	62.38	46.88	41.40	5.53	6.19	52.41	53.07	89.44	13.00	0.77
Age 0-15 and <200% Poverty	13.23	69.07	44.06	46.09	5.07	4.79	49.13	48.84	89.69	9.41	0.80

Notes: Entries are percentages of person-months for individuals with valid SSN who are living in California in the month.

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
<i>By Age:</i>	
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
<i>By Ethnicity:</i>	
White Non-Hispanics	85.7
Black Non-Hispanics	76.6
Asian Non-Hispanics	81.6
Hispanics	76.4
<i>By Poverty Status and Age:</i>	
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
<i>By Reported Medicaid Coverage:</i>	
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 <i>and Living in California</i>	25.32	26.95	20.09
Percent of Valid Months in California	96.79	96.56	97.62
<b>Person Months:</b>			
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
<b>Addendum:</b>	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: AFDC and Medi-Cal Enrollment, 1988-1996

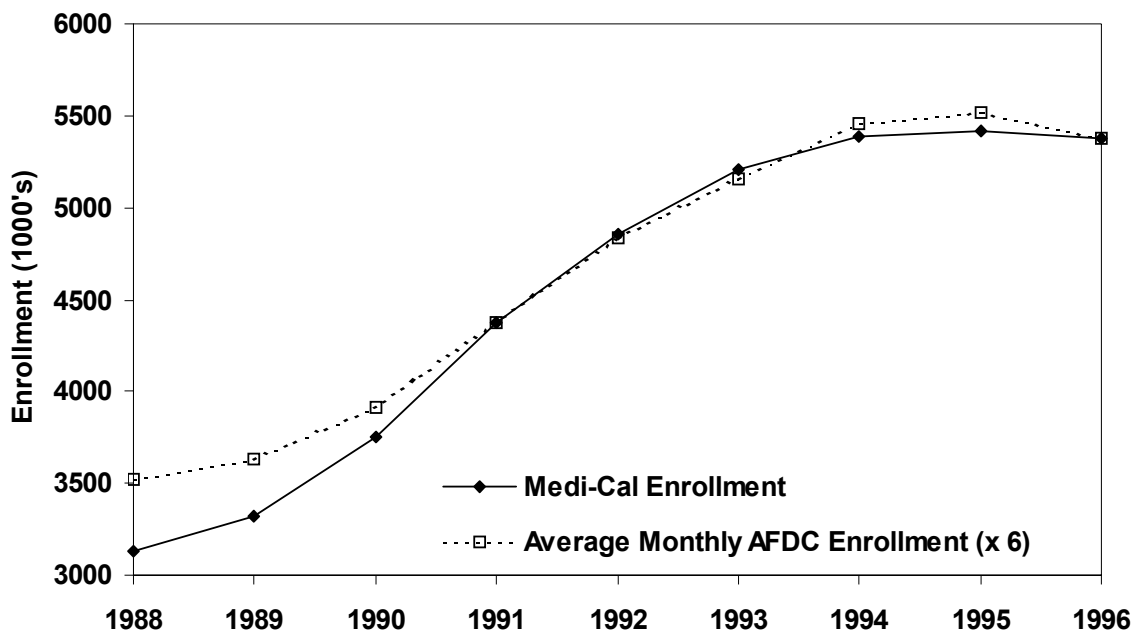


Figure 2: Medicaid Participation Rates in California, SIPP versus Administrative Data

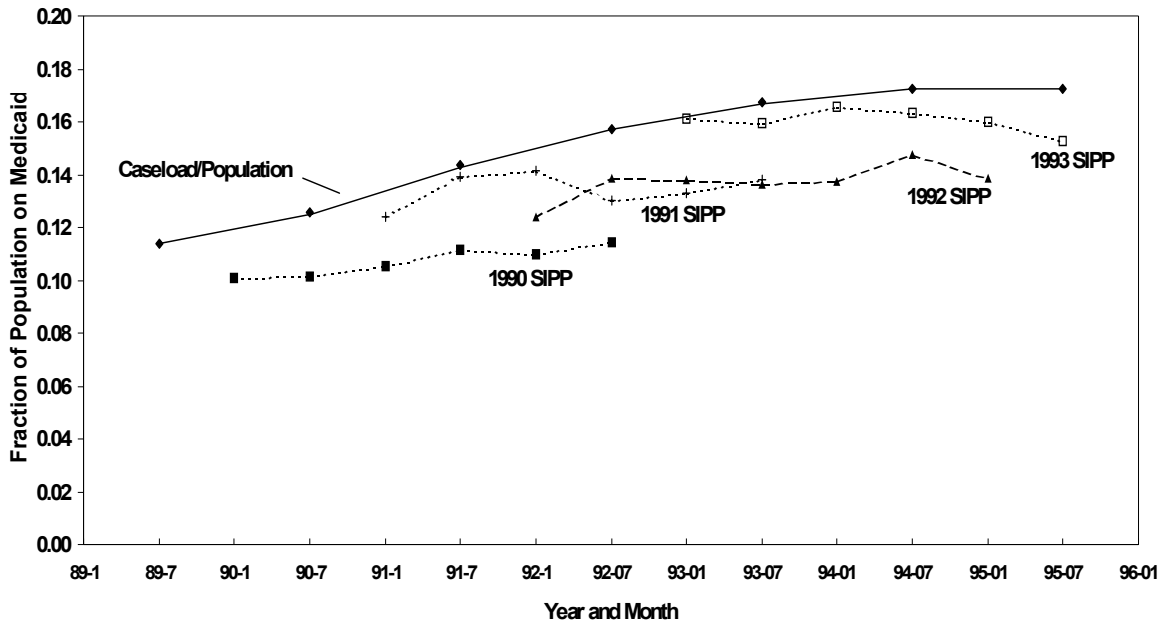


Figure 3: Medicaid Coverage Rates for SIPP California Sample, by Family Income

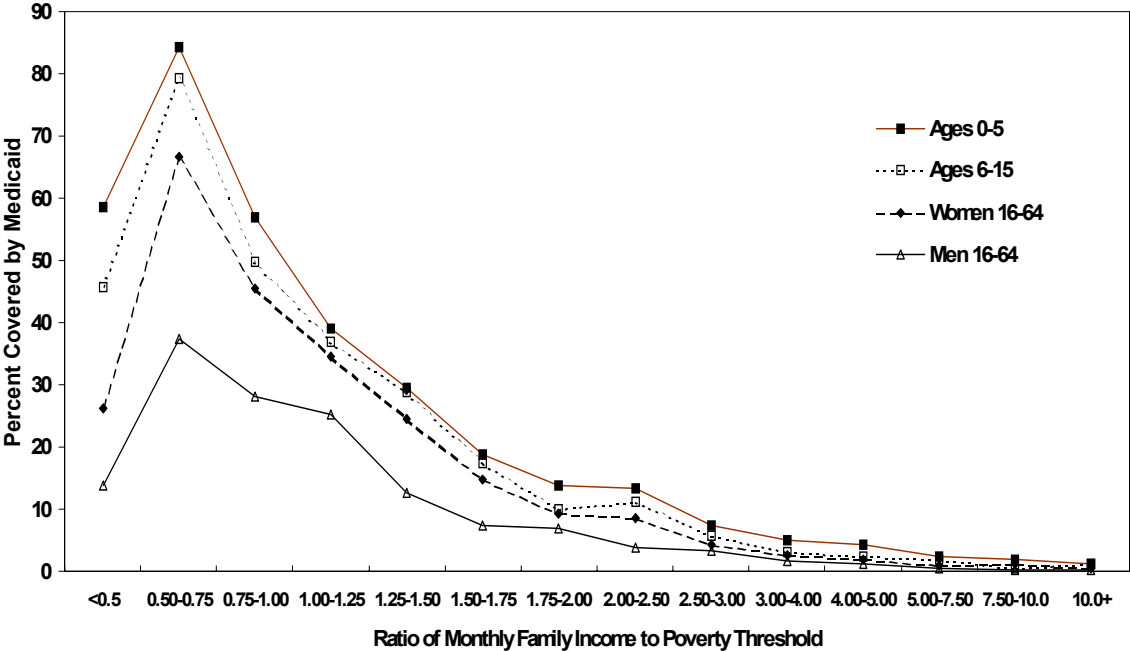


Figure 4: Medicaid Coverage, Entry and Exit Rates by Survey Month

