

**Research Project to Understand the Medicaid Undercount:
The University of Minnesota's State Health Access Data
Assistance Center, the Centers for Medicare and Medicaid
Services, the Department of Health and Human Services
Assistant Secretary for Planning and Evaluation, the National
Center for Health Statistics, and the U.S. Census Bureau**

***Phase III Research Results:
Refinement in the Analysis of Examining Discrepancies between
the National Medicaid Statistical Information System (MSIS)
and the Current Population Survey (CPS) Annual Social and
Economic Supplement (ASEC)***

Final
July 14, 2008

Note: This paper was made possible by grant 052084 from the Robert Wood Johnson Foundation to the State Health Access Data Assistance Center (Michael Davern, PI) with additional support supplied by the Office of the Assistant Secretary for Planning and Evaluation (ASPE), the Centers for Medicare and Medicare Services (CMS), the National Center for Health Statistics (NCHS), and the U.S. Census Bureau. This paper has undergone a limited review by all the participating organizations in accordance to existing agreements among these organizations. The views expressed are those of the author(s) and not necessarily those of ASPE, NCHS, CMS, or the US Census Bureau.

Contents

Executive Summary	ii
I. INTRODUCTION.....	1
II. OBJECTIVE AND SCOPE.....	1
III. RESULTS OF THE ANALYSIS	2
Accounting for Phase II Unexplained Discrepancy:.....	2
Analysis of Person-Identification Procedures Enabling Phase II Matching	12
IV. CONCLUSION	15
V. APPENDICES	17

I. Introduction

This paper describes the results of the third phase of a multi-phase⁴ research project of University of Minnesota's State Health Access Data Assistance Center (SHADAC), Centers for Medicare and Medicaid Services (CMS), Assistant Secretary for Planning and Evaluation (ASPE), National Center for Health Statistics (NCHS), and the U.S. Census Bureau. The research is designed to explain why discrepancies exist between survey estimates of enrollment in Medicaid and the number of enrollees reported in state and national administrative data.

The research done for this project includes both national and state-level analysis. National files include the Medicaid Statistical Information System (MSIS), the Medicaid Analytic eXtract (MAX), and the Medicare Enrollment Database (EDB). Survey files include the Current Population Survey/Annual Social and Economic Supplement (CPS/ASEC) and the National Health Interview Survey (NHIS). Additionally, the following states participated in this study: Florida, Maryland, California, New Jersey, Louisiana, and Minnesota. Participating states provided data from their Medicaid and State Children's Health Insurance Program (SCHIP) enrollment files to CMS, and CMS in turn provided these data to the U.S. Census Bureau.

Understanding differences between enrollment data and survey data will benefit the Census Bureau and other participating agencies by suggesting possible improvements to CPS and other surveys. It will also engender a better understanding of existing CPS insurance data that provide a more accurate view of Medicaid and other insurance coverage, for U.S. residents. As such, this research will enable a revised computation of the number of U.S. residents without health insurance. We consider the results presented in this paper to form only a basis for making such an analysis. Without a sophisticated strategy to reassign surveyed persons categorized by CPS as uninsured but established by this study to have Medicaid coverage, an estimate of the uninsured would likely be inaccurate. We expect additional work released by team members to speak directly to this issue.⁵

II. Objective and Scope

Phase III consists of an attempt to further vet and refine the accounting for the MSIS record count-CPS estimate discrepancy on the number of persons enrolled in Medicaid in the year previous to CPS/ASEC interview. Toward this goal, the Phase III analysis as it is presented here consists of two phases:

- Attempt to account for the count-estimate discrepancy left unexplained by the Phase II analysis, including a re-casting of the Phase II presentation of that discrepancy.
- Analysis of the quality of the person identification procedures used to enable the Phase II match of CPS to MSIS.

⁴ Description of the first four phases can be found in Appendix I

⁵ Jacob Klerman of Abt Associates Inc. is performing an analysis that will adjust CPS estimates of the number of uninsured by adjusting for MSIS-identified Medicaid enrollees reported uninsured in a way that accounts also for false-positive Medicaid reporting, incomplete matching, and overlapping reported and actual types of health coverage.

III. Results of the Analysis

Accounting for Phase II Unexplained Discrepancy:

Universe coverage differences between the CPS and MSIS

Phase III of the SHADAC project includes an attempt to determine the number of people covered by Medicaid according to MSIS who were living in institutions at the time of the March 2001 CPS interview. People living in institutions, in military facilities, overseas, or who have died before the interview are not eligible for the CPS, while MSIS covers all Medicaid enrollees regardless of living situation. Because of this difference in universe coverage, we would like to know how much of the Medicaid population is outside the scope of the CPS. This will help explain the CPS's apparent underestimate of the number of Medicaid enrollees generally, and in particular the 2.8 million persons missing from the CPS estimate left unexplained by the Phase II analysis.

We are particularly interested in the institutional population. Decedents – those who were living in the year of interest but died before the interview the following year – were simple to remove from the analysis in Phase II, and the other populations omitted from the CPS should be very small. Though it may be non-trivial in number, we believe only a relatively small subset of the Medicaid population left the U.S. before the CPS interview, perhaps limited principally to some migrant workers in the Southwest. Similarly, there should only be a small number of military personnel who were on Medicaid at some time during the year of interest before enlisting in the armed forces.⁶

Methods of determining the institutional status of MSIS persons

Phase II attempted to remove the institutional population from the comparisons by looking for MSIS persons in the Master Address File Auxiliary Reference File (MAF-ARF). This file is a compendium of person (identified by PIK)-address (identified by MAF-ID) pairs. Each pair derives from an administrative record showing a specific person associated with a particular address, usually a residence. The MAF-ID references a specific MAF record that, among myriad data elements, contains information about the type of living quarters, including whether the address is a group quarter, either institutional or non-institutional. This method – finding MSIS persons in the MAF-ARF and their addresses in the MAF – determined that 209,850 MSIS persons had at least 1 institutional address.⁷

As explained in the Phase II report, there are several reasons to doubt the comprehensiveness of this count. First, many MSIS persons cannot be found in the MAF-ARF. Second, a person living

⁶ About 3,900 MSIS persons according to the analysis that follows in this report.

⁷ Among identified MSIS persons, only 87.5% were found in MAF-ARF for Year 2000.

in an institution may be shown at the family's private residential addresses on administrative forms instead of the address of the institution. Finally, some addresses in the MAF may not be properly coded as institutions. These limitations will tend to cause underestimation, and motivate further work to determine the institutional status of MSIS persons.

Phase III employs a multi-step procedure to elucidate the discrepancy exposed by Phase II research, exploiting Census 2000, MAX, and State Medicaid data. Each step proceeds only with the residual persons of the preceding step, that is, any person found in Census 2000 in Step 1 is not subject to Step 2 or Step 3. Any person eligible for assignment to institutional living in Step 2 is not subject to the Step 3a match to the MAF, and any person matched to the MAF in Step 3a does not proceed to clerical inspection in Step 3b.

Step 1 – Census 2000 Search. We match MSIS persons to Census 2000 data by PIK.⁸ All places of residence in Census 2000 have known institutional status based on fieldwork immediately preceding the Census.

Step 2 – MAX Search. We use information in the MAX files indicating the number of days Medicaid paid for institutional living. Proportions generated by a cross-tabulation with Census 2000 provide some level of temporal and conceptual consistency with Census 2000, and thereby with the CPS. Note that the proportions are developed from persons *for whom Medicaid is paying for institutional care* who **are** found in Census 2000 data and is applied only to persons *for whom Medicaid is paying for institutional care* who **are not** found in Census 2000.

Steps 3a and 3b – Clerical Inspection. We match the addresses in the files supplied by a few states to the MAF. Then we clerically inspect and search on the internet for a sample of the addresses not matched, and adjust the U.S. estimate accordingly.

The results of Steps 1 and 2 provide data for regenerating the Phase II analytic tables omitting more of the institutional population than was possible in Phase II. Step 3, because it relies on data available only in a subset of states, cannot readily be used to revise Phase II tables.

The discussion below follows the Phase III Table 1 presentation closely. It discusses an analysis that includes only MSIS persons identified with a PIK, considered on an unduplicated basis, who received full Medicaid benefits sometime in calendar year 2000, and were still living at the time households were interviewed in March 2001.

⁸ PIKs are determined according to processes described in the Phase II report.

Phase III Table 1. Estimate of MSIS Persons Excluded from the CPS Universe, Based on Census 2000, MAX, and State Medicaid Data

	CA	FL	LA	MD	MN	NJ	6 States	Other States	All States
a Total Non-Institutional Population (CPS Control Totals)	34,000,000	16,020,000	4,370,000	5,260,000	4,890,000	8,390,000	72,950,000	206,550,000	279,500,000
b MSIS Total (SSN-identified Medicaid Enrollees Only)	5,320,000	1,851,000	659,000	464,000	528,000	678,000	9,500,000	26,700,000	36,200,000
c CPS/ASEC Augmented CAID Estimate	4,260,000	1,340,000	500,000	240,000	340,000	540,000	7,230,000	18,840,000	26,070,000
d <i>Original Discrepancy (b - c)</i>	1,060,000	511,000	159,000	224,000	188,000	138,000	2,270,000	7,860,000	10,130,000
e CPS/ASEC Corrected-Enrollment Response Estimate	4,800,000	1,780,000	660,000	340,000	400,000	640,000	8,620,000	24,830,000	33,450,000
f <i>Unexplained Discrepancy (b - e)</i>	520,000	71,000	-1,000	124,000	128,000	38,000	880,000	1,870,000	2,750,000
1 Census Institutional	66,000	31,700	17,550	11,750	18,750	16,800	162,550	602,000	765,000
2 MAX at least 1 week	20,200	11,200	7,550	4,550	4,350	7,350	55,200	178,500	233,700
3 MAX proportion	13,050	7,500	4,800	3,150	3,100	5,050	36,650	120,600	157,300
4 Census Institutional + MAX proportion	79,050	39,200	22,350	14,900	21,850	21,850	199,200	723,000	922,000
<u>Added by State Medicaid Addresses matched to MAF</u>									
5 ALL of a person's addresses are MAF institutional	2,600	4,150	150	1,350	200	450	8,900		
6 Ratio (percent) line 5 / line 4	3.3	10.6	0.7	9.1	0.9	2.1	4.5		963,000 *
<u>Added by Clerical Inspection of State Medicaid Addresses</u>									
7 Street addresses deemed institutional or out of scope							1,900		
8 PO Box addresses deemed institutional or out of scope							10,000		
9 Total addresses deemed institutional or out of scope							11,900		
10 Ratio (percent) line 9 / sum of line 4 + line 5							5.7		1,018,000
11 Institutional, Phase II match to STARS	10,500	8,700	2,600	2,100	2,250	6,600	32,750	123,600	156,400
12 TOTAL out of scope							220,100	955,000	1,175,000
13 MSIS in scope of CPS							9,280,000	25,760,000	35,050,000

Note: this analysis was not conducted in a way that allows the development of state-specific estimates

14 Note: Table includes MSIS people with PIK, unduplicated, with full Medicaid benefits, living at the time of the survey

15 *Note: This value is generated by incrementing Row 6 (All State) value by the 6-State (mean) ratio (4.5%)

Details and results

Step 1. A relatively straightforward way to determine if an individual was living in institutional group quarters in 2000 is through Census 2000 data. Field operations of the Decennial Census list every living quarter as residential, non-institutional group quarters, or institutional group quarters. This data set, with 85% of Census-recorded individuals identifiable by a PIK, can supply information on the living situation of most persons in the MSIS file who are also identified by the PIK. Although the reference date of Census 2000 is April 1, 2000, about a year before the CPS data was collected, we can assume that someone's 2001 status should not diverge substantially from what we find in Census 2000, because institutional living is typically long-term. Moreover, the aggregate count and characteristics of institutionalized persons likely change little from year to year.

Line 1 of Table 1 shows that Census 2000 indicates that 765,000 full-benefit enrollees were living in institutions. Census 2000 uses the same definitions of residence type as the CPS. The types of institutional and non-institutional group quarters are as follows:

Institutional Group Quarters

- Correctional institutions
- Juvenile institutions
- Nursing homes
- Hospitals/schools for the handicapped

Non-Institutional Group Quarters

- College quarters
- Military quarters
- Emergency and transitional shelters
- Shelters for runaway, neglected, and homeless children
- Shelters for abused women
- Soup Kitchens
- Regularly scheduled mobile food vans
- Targeted non-sheltered outdoor locations
- Group homes/halfway houses
- Crews of maritime vessels
- Agriculture workers dormitories
- Other workers dormitories
- Job corps and vocational training facilities
- Staff residents of military institutions
- Staff residents of non-military institutions
- Religious group quarters
- Other non-household living situations
- Living quarters for victims of natural disasters
- Resident care facilities providing protective oversight

Step 2. A substantial number of MSIS persons (about 30%) could not be found in the Census 2000 data and require further work to determine their type of residence. The failure to find someone in Census 2000 occurs for two main reasons:

1. We cannot identify every individual in Census 2000 with a PIK, because the name, date-of-birth, or address information collected for Census 2000 does not allow a definite determination.
2. Not all MSIS persons are in the Census 2000 data, either because they were missed by Census operations or they were not living in the United States on Census Day.

For these individuals, the MAX data set provides useful supplementary information. It indicates how many days Medicaid paid for an individual's stay in an institutional facility. These facilities include:

- nursing facilities
- intermediate care facilities for the mentally retarded
- long term care mental hospitals for the aged
- long term care inpatient psychiatric facilities (age < 21)

Persons not found in Census 2000 and for whom Medicaid payments covered at least one week in one of these facilities amount to an additional 233,700 individuals possibly out-of-scope for CPS (line 2), an upper bound of supplementary estimates possible from the MAX data. However, many of these individuals may have lived in the facility only temporarily and otherwise at a residence included in the CPS universe, so the upper bound of 233,700 may be too high.

Apart from the problem of timing, which we can at least partially deal with using the number of weeks indicated by the MAX, the definition of "institution" is a problem. That is, the MAX facilities may not conform to the Census Bureau's concept of institutional group quarters. Altogether, we consider the Census 2000 determination of residency type more reliable⁹ than MAX information. For this reason, we built a simple model that imputes Census residence status from MAX institutionalization status. To do this, we calculated the proportion of individuals in each MAX institution who were living in institutional group quarters in Census 2000, by age and number of weeks, among those persons found in Census 2000. A summary of the computations of these proportions is shown below in Table 2. In Table 2, the Denominator of the Rate is the number of persons found in both MAX and Census 2000 who are in each demographic category. For example, the upper right table value, 400, represents MAX persons found in Census 2000 data who were age 0 - 17 and had during Year 2000 35 to 52 weeks paid in an intermediate care facility for the mentally retarded in one of the six data-providing states. Among those 400 persons, 34.5% were recorded in Census 2000 as living in institutional group quarters.

Applying these proportions to individuals in the MAX institutions but not found in Census 2000 should produce a reasonable estimate of the true institutional status and characteristics of persons not found in Census 2000 who had weeks in MAX facilities paid for by Medicaid. This

⁹ In the sense of being more conforming to the type of residence determinations made for CPS.

procedure generates an estimate of 157,300 full-benefit enrollees who are out-of-scope for CPS, as shown in line 3 of Table 1. This proportional method supplementing the direct Census 2000 determination of residence status (in Step 1, which is only possible for those persons identified in Census 2000) brings the estimate of total institutional MSIS population to 922,000 (line 4).

The calculated proportions of full benefit enrollees living in institutions vary widely by type of facility, number of weeks, and age, as shown in Table 2. MAX nursing facilities seem to align best with the Census 2000 data, and children's inpatient psychiatric facilities the worst. Perhaps families tend to include these children in the household roster when responding to a census or survey. In this case the concept of "institution" per se does not cause the classification difference, but rather the idea of where someone resides.

Phase III Table 2. Percent Rates of Census 2000 Institutionalization Among MAX Institutional Population, by Type of MAX Facility, State Grouping, Number of Weeks Paid for by Medicaid, and Age Category

	<u>Weeks Paid for by Medicaid</u>			<u>Denominator of Rate</u>		
	<u>1 - 17</u>	<u>18 - 34</u>	<u>35 - 52</u>	<u>1 - 17</u>	<u>18 - 34</u>	<u>35 - 52</u>
<i>Intermediate Care Facility for the Mentally Retarded, Selected States</i>						
<u>Age</u>						
00 - 17	12.4	—	34.5	100	—	400
18 - 64	41.8	43.9	37.1	950	1,250	13,400
65+	—	52.8	35.8	—	100	750
<i>Intermediate Care Facility for the Mentally Retarded, Other States</i>						
<u>Age</u>						
00 - 17	10.7	33.8	52.9	250	200	1,250
18 - 64	22.1	33.4	40.9	1,900	2,600	48,350
65+	38.6	49.2	44.5	150	200	4,250
<i>Long Term Care Mental Hospital for the Aged, All States</i>						
<u>Age</u>						
00 - 17	17.8	—	—	500	—	—
18 - 64	12.9	—	—	3,000	—	—
65+	29.4	82.0	95.0	1,350	300	1,050
<i>Long Term Care Inpatient Psychiatric Facility (Age < 21), All States</i>						
<u>Age</u>						
00 - 17	5.7	20.0	38.4	11,200	1,250	700
18 - 64	15.6	46.1	66.8	3,550	300	200
65+	—	—	—	—	—	—
<i>Nursing Facility, All States</i>						
<u>Age</u>						
00 - 17	12.0	34.4	75.1	450	150	750
18 - 64	21.5	52.3	91.5	26,200	12,000	66,350
65+	24.3	49.1	94.3	82,250	61,300	410,400

Notes: “—” denotes less than the minimum necessary to meet disclosure limits.

Table includes MSIS people with PIK, unduplicated, full Medicaid benefits, not deceased, at least 1 week in a MAX facility paid for by Medicaid, matched to Census 2000. Selected states are CA, FL, LA, MD, MN, and NJ.

Where do persons whom the MAX lists as institutional live according to Census 2000, if not in an institution? Those whom we find in Census 2000 mostly live in group homes and halfway

houses (68.8%), with the rest mostly split about evenly between residential housing (13.7%) and resident care facilities providing protective oversight (13.4%). It would seem that the largest part of the difference in the definition of “institution” stems from the existence of community care.

Table 3 delineates the distribution of the 922,000 institutional MSIS persons in line 4 by age and state. The institutional MSIS population is mostly 65 years old or older, with about 30% in the range of 18 to 64 and 5% in the range of 0 to 17. In California, it is somewhat younger, with 42.0% between 18 and 64 years old and only 52.5% aged 65 or older. Minnesota is also exceptional, with a somewhat older institutional MSIS population; 76.3% are 65 years old or older, and only 19.2% in the range of 18 to 64.

Phase III Table 3. Percent Distribution by Age of Institutionalized MSIS Population Under Definition "Census Institutional + MAX Proportion" by Age

	<u>CA</u>	<u>FL</u>	<u>LA</u>	<u>MD</u>	<u>MN</u>	<u>NJ</u>	<u>Other</u>	<u>Total</u>
<i>Computed Institutional Population</i> →	79,050	39,200	22,350	14,900	21,850	21,850	723,000	922,261
<u>Age</u>								
00 – 17	5.6%	3.6%	4.1%	5.5%	4.6%	3.6%	4.8%	4.8%
00 – 05	0.3%	0.3%	0.3%	0.1%	0.1%	0.2%	0.2%	0.2%
06 – 14	1.2%	1.1%	1.0%	1.5%	1.3%	1.2%	1.4%	1.4%
15 – 17	4.1%	2.2%	2.8%	3.9%	3.2%	2.2%	3.2%	3.2%
18 – 64	42.0%	28.6%	32.1%	30.0%	19.2%	26.0%	28.7%	29.7%
18 – 44	25.7%	16.4%	16.6%	16.4%	10.1%	13.0%	15.8%	16.5%
45 – 64	16.3%	12.2%	15.5%	13.6%	9.1%	13.0%	12.9%	13.2%
65+	52.5%	67.8%	63.8%	64.5%	76.3%	70.4%	66.5%	65.5%
Total¹⁰	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Step 3. The possibility of someone residing in a facility without payment by Medicaid motivates this step, which examines the addresses of persons in the State Medicaid data supplied by 6 states (7 states provided data, but one was determined unusable due to incompleteness). In this step we only look at CY 2000 full benefit enrollees who had not been determined to be institutionalized in Phase II, were not found in Census 2000, and did not have any weeks in a facility paid for by Medicaid.

First, in Step 3a, matching the addresses for these persons to the MAF (which indicates the type of living quarters) provides information about persons who were not found in Census 2000 and for whom MAX shows no institutional Medicaid care. Because someone can have more than one address in the State Medicaid data, we assigned institutional status by requiring all of a person's

¹⁰ Note: Details may not sum to totals because of rounding.

addresses to be institutional according to the MAF. This rule inflated the U.S. estimate of institutionalized Medicaid enrollees by 4.5 percent to 963,000 (line 6).

Finally, Step 3b investigates persons not found in Census 2000, who had no weeks in a MAX facility, and who had at least one address in the State Medicaid data that did not match the MAF in Step 3a. Searching for information about a sample of these addresses on the internet might identify more full-benefit enrollees living in institutions. To deal with the problem of multiple addresses, we created a weight for each person equal to the inverse of the number of addresses recorded for a person. This weight is akin to a probability of residing at the address on a particular day. We also created a count of persons at each address, because a large number of enrollees at an address might indicate a group quarter and perhaps an institution.

Typically an address fails to match the MAF because it is new construction not yet captured by MAF updating procedures, it is a Post Office box or commercial address, or it is garbled. Here, we are interested in the 300,136 persons not found in Census 2000, without weeks paid for in a MAX facility, and who had at least one address that did not match the MAF. The state Medicaid addresses of this group follow the typical pattern; on a weighted basis (where weights equal the inverse number of addresses for a person) about 70 percent of the addresses are post office boxes. Most of the remaining addresses are street addresses, and at 17,747 of these addresses reside five or more Medicaid enrollees. We sampled 100 of these “clustered” street addresses for clerical inspection and internet search, and categorized persons at the sampled addresses as follows:

<u>Percentage of Clustered Address (Computed After Weighting*)</u>	<u>Address Type</u>
26.3%	Residential
1.1%	Group Quarters (GQ) Unknown whether institutional
0.4%	Group Quarters (GQ) Non-institutional
5.2%	Group Quarters (GQ) Institutional
3.9%	Commercial
1.6%	Other Out-of-scope
26.1%	Unknown (at least 1 hit on the Web but unknown category)
29.6%	Not Found (no hits on the Web)
5.8%	<u>Insufficient Information (garbled)</u>
100.0%	Total

*Note: For each sampled person, weight equals product of the inverse of the probability of selection for him or her (which is uniformly $100 / 300,136 \approx 0.0033$) and inverse of number of different address recorded for him or her.

Deemed outside the CPS universe, the categories GQ Institutional, Commercial, and Other Out-of-scope totaled 10.6 percent of persons in the street address sample. This figure produces an estimate of 1,900 MSIS persons (line 7). The addresses of these persons included several interesting living situations such as an adoption agency (which presumably acquires Medicaid coverage for children awaiting adoption), a migrant workers’ camp, some social service agencies, and “no permanent address.”

We also investigated the Post Office box addresses. There are 210,070 persons at these addresses total, and at 71,025 persons' addresses, five or more Medicaid enrollees receive their mail. We sampled 94 of these "clustered" P.O. box addresses, performed an internet search, and found 14.1 percent of persons using commercial P.O. boxes, all but one of them social service agencies; 5.2 percent non-institutional group quarters; 2.1 percent residential; and the remaining 78.6 percent unknown or not found. The figure for commercial P.O. boxes, 14.1 percent, produces an estimate of 10,000 out-of-scope MSIS people (line 8).

Line 9 contains the total out-of-scope population estimated by the street address and P.O. box samples, 11,900 persons. This represents an adjustment of 5.7 percent and brings the U.S. estimate to 1,018,000 (line 10).

Impact of Phase III Analysis on Phase II Results

The Phase II report left a gap of 2.8 million Medicaid enrollees unexplained. Here Phase III lowers this figure by 1.0 million to 1.8 million. In other words, 36.7% of the unexplained Phase II discrepancy can now be explained as MSIS-determined enrollees living in institutions who were not identified as such in the Phase II analysis. The remaining discrepancy may be explained in part by sample variance. However, since it is highly unlikely that sample variance accounts for this entire amount, there is still an unexplained discrepancy.

Since the Phase III analysis was able to determine additional likely out-of-scope persons due to institutional residence, we thought it would be useful to recompute Phase II's accounting of the discrepancy removing these persons. The persons who are removed are either those identified as living in group quarters in Census 2000 or who were imputed as such as part of Phase III, Step 2. As it should be clear that these persons so identified may not in all cases have been truly out-of-scope (by which we mean institutional residents at the time of the CPS-ASEC 2001 interviews), the revised tables that we produce should be considered non-authoritative. Nevertheless, since we believe Phase III, Steps 1 and 2, are largely correct in aggregate, even if somewhat less so at the individual level, the revised tables show how the Phase III explained discrepancy differentially affects the count-estimate discrepancy among various demographic categories. Revisions to Phase II tables based on Phase III results are provided in Appendix II to this report.

Limitations

As mentioned previously, the Phase III discrepancy analysis depends heavily on Census 2000 data which was collected a year before the CPS interview and refers to April 1, 2000; and MAX 2000 data which refers to calendar year 2000. We chose to use 2000 data because the coverage of persons and accuracy of institutional coding in Census 2000 is better than any other data source insofar as the definition of an institution is the same as in the CPS, and because we are aiming at a point-in-time estimate of the institutional population to parallel the cross-sectional nature of the CPS. Moreover, the size and characteristics of this population likely change little from one year to the next.

As mentioned in the Phase II report, a more important limitation of this analysis is the exclusion of MSIS persons who lack a PIK. Another is Step 2's assumption that within age, weeks, and facilities categories, MSIS persons not found in Census 2000 have the same rate of institutional living as those found in Census 2000. This is unlikely to be the case; however, how the rates may differ is not known.

A minor limitation is dependence on the MAF. The quality of the MAF's updating procedures is unknown. Misclassifications of group quarters may occur in cases where addresses have converted recently between commercial, residential, non-institutional, and institutional uses. However, the MAF is the best source of information available about addresses in the U.S. that conforms to the Census Bureau definition of an institution. Another minor limitation is the adjustment calculated in Step 3b, which may be too large if a substantial number of MSIS persons use addresses of social service agencies *in addition* to residential addresses, and not because of homelessness, long-term transience, or actually dwelling there. The weighting and calculation procedure applied to the person-address pairs assumes that each person has only one address at a time, and having multiple addresses would cause in-scope persons to be counted partially as out-of-scope, biasing the adjustment upward. On the other hand, we did not have time to clerically inspect unclustered addresses, or determine the status of persons with bad addresses – omissions which work in the opposite direction and may make the adjustment too small.

Analysis of Person-Identification Procedures Enabling Phase II Matching

For the Phase III analysis, seven states provided extracts of their Medicaid program rolls intended to cover the period of years 2000 to 2002. The primary advantage of these data relative to MSIS data is that they included the names and addresses of the enrollees. It seems reasonable to expect that having this information would allow for better confirmation and determination of person identification than was possible without it. For this reason, we were able to use these data to evaluate the quality of the validation of person identity that was made from MSIS data alone and that was the basis for the survey-administrative record match analysis that comprised Phase

II.

The person identification procedure used with MSIS for Phase II consisted of a verification analysis of the Social Security Number (SSN) provided on each MSIS record. This analysis compared the date of birth and sex on the MSIS record to that recorded by the Social Security Administration (SSA) in the processing of that person's Social Security Number. Since it is rather unlikely that date of birth would be similar for an incorrectly provided SSN, we used near similarity of the MSIS date of birth and the SSA Numerical Identification File (Numident)¹¹ date of birth as well as agreement on sex to validate the provided SSN. Details of this process are shown in Appendix III.

To evaluate the quality of this validation, we used the Data Integration Division's Person Identification Validation System (PVS) to determine independently the identity of Medicaid enrollees for the states that provided this data (collectively called the SNACC Extract) and compared the results of these determinations (special verification vs. PVS verification). For this purpose, the PVS process consisted of a routine validation step that compared Medicaid program collected data for a person to data collected for the person in SSN processing. The types of data used for this validation were names, dates-of-birth, and sex. In cases where the verification could not be made, we conducted a name search that looked for the enrollees by name in the Census-summarized Social Security registration history file (called Census Numident) that also compared the similarity of dates-of-birth and sex.

Table 4 presents a detailed review of the comparison of the special verification, on which Phase II analysis is based and that used only date-of-birth and sex to verify record holders' identities, with the PVS verification, that, in addition, also uses names. Because three years of MSIS are included that were each processed independently, it is possible for SSNs or dates-of-birth or sex to be different on various years' records for the same client account, in which case it is possible for that client account to have a verified identity and a non-verified identity, even if as they should be, both SSNs are the same. More commonly, an MSIS account either had identity verified or not verified consistently.

¹¹ The Numident is a comprehensive file of Social Security program transaction records indexed by Social Security Number.

Phase III Table 4. Joint Distribution of Special Verification Status and PVS Verification Status

**Joint Distribution of Special Verification Status and PVS Verification Status
For Accounts Included in MSIS and SNACC Extract**

Row #	Special Verification Status	PVS Verification Status	Same SSN on MSIS and State Extract	PVS SSN Matches Original SSN	Count	Percentage of Total
1	Not Verified	Not Verified	-	Yes	570	0.0%
2	Not Verified	Not Verified	No	No	80	0.0%
3	Not Verified	Not Verified	Yes	No	36,060	0.6%
4	Not Verified	Found in Search	-	No	240	0.0%
5	Not Verified	Found in Search	No	No	40	0.0%
6	Not Verified	Found in Search	Yes	No	48,300	0.8%
7	Not Verified	Found in Search	Yes	Yes	10	0.0%
8	Not Verified	Verified	-	No	10,260	0.2%
9	Not Verified	Verified	No	No	2,230	0.0%
10	Not Verified	Verified	Yes	Yes	20,160	0.3%
11	Verified	Not Verified	No	No	90	0.0%
12	Verified	Not Verified	Yes	No	10,210	0.2%
13	Verified	Found in Search	No	No	20	0.0%
14	Verified	Found in Search	No	Yes	220	0.0%
15	Verified	Found in Search	Yes	No	4,200	0.1%
16	Verified	Found in Search	Yes	Yes	1,480	0.0%
17	Verified	Verified	No	No	1,110	0.0%
18	Verified	Verified	Yes	Yes	<u>6,245,160</u>	<u>97.9%</u>
					6,380,440	100.0%

For the PVS process, identification can be confirmed in the validation step, or failing that, from a successful search using the name values provided on the state extract against the Census Numident. The table also shows whether the provided SSN was consistent for the same client account (as identified by a unique MSIS-ID) between MSIS and the state extract data. The remaining classification variable shows whether the PVS-verified SSN matches the MSIS-originally provided SSN. A count (and percentage of total) is then shown for each crossing of the descriptive variables.

A review of this table shows that most client records had their MSIS SSN verified and the verified SSN was confirmed by the PVS run against the state-provided Medicaid extract data. Among the 6,262,490 client accounts with specially verified person identity, only 10,210 + 4,200 = 14,410 (0.2%) of these verifications were repudiated by PVS processing. Also, PVS processing, which was able to use names for search, was able to establish person-identity for an additional 81,240 (Table 4 Rows 4 – 10). This represents a 1.3%¹² addition over confirmed

¹² From Table 4: Numerator Rows: 4 – 10 = 81,240; Denominator Rows: 14, 16, and 18 = 6,246,860.

special verified client accounts – not enough, we think, to noticeably affect a combined Phase II-type analysis for states included in this analysis.

IV. Conclusion

The Phase III analysis is intended to qualify the results obtained in Phase II. The larger task is to account for the existing discrepancy between MSIS enrollee counts and CPS enrollee estimates after correcting for known frame issues and survey mis-reports. Starting with an unexplained discrepancy of about 2.8 million enrollees for calendar year 2000, the analysis is able to show that about an additional 1.0 million of the persons counted from MSIS as enrolled were likely out-of-scope of the CPS universe because they were either residing in residential institutions (such as nursing homes) or because they were homeless. This 1.0 million figure should be considered as a minimum bound on the number of additional MSIS persons who were out-of-scope. Note that several factors support this contention:

- For this analysis, non-Census 2000-identified enrollees were imputed to institutional group quarters at the rate found for Census 2000-identified enrollees. However, since the non-identified persons are less functional than usual, it seems not unreasonable to assume they are out-of-scope at a higher rate.
- For enrollees not identified receiving long-term care services in MAX, out-of-scope rates were estimated based on a review of a sample of addresses. However, it is not clear that our clerical review was able to identify all institutional addresses. In addition, homeless persons or persons not considered permanent members of a (private) household may have a residential address or non-institutional post office box recorded for them in Medicaid administrative records.

Even if all out-of-scope enrollees were accounted for, likely there would still be a remaining discrepancy between the MSIS count and the CPS estimate. This may relate to several factors such as:

- The observed occurrence of households not rostering all household members, particularly young children, for surveys.
- Potentially greater response rates by higher income families compared to lower income families, which may be insufficiently controlled by non-response adjustments. This would represent the population with greater than actual income.
- That some families reside in hidden units, particularly informally subdivided multi-family structures that are not in the CPS address frame and may not be captured by CPS/ASEC multi-unit listing procedures. Likely this happens at a higher rate for lower income families.
- That some families in non-standard housing may have intentionally provided non-existent addresses on their Medicaid application.

Even so, that we are able to account for 33.4 million of 35.2 million enrollees evidences the high level of performance of CPS for this hard-to-reach population. The CPS includes low-income households in the survey frame, produces completed interviews for them, and then accounts for

non-respondents through weighting computations. Our ability to account for 33.4 million of 35.2 million enrollees reiterates the importance of the relatively high number of MSIS indicated-Medicaid enrollees that were reported to be uninsured (false-negatives) on the CPS-derived estimates of Medicaid enrollees and suggests possible biases of other survey estimates of low-income support program participants.

The Phase III analysis also included a review of the person identification procedures used in the Phase II analysis. For the Phase II analysis, verification of person identity was limited to validation of MSIS-provided Social Security Numbers (SSNs) using MSIS-recorded date-of-birth and sex values. Because this is less stringent than the Census Bureau's standard procedure used for handling administrative records (as was necessary because the data did not include names or addresses), it was thought valuable to see how these validations compared to what would have resulted had names and addresses been available. We were able to make this comparison because several states provided this additional identification data for their Medicaid enrollees.

This comparison confirmed 99.8%¹³ of the original validations. Also, the additional identifications that could be made only exceeded the original amount by about 1.4%¹⁴. We conclude then that our comparison of the modified validation procedure used for Phase II against the more extensive standard showed that the modified procedure performed quite well and that limitations of this modified procedure were unlikely to materially affect the analysis conducted in Phase II.

¹³ From Table 4: Numerator: Rows 14, 16, and 18 = 6,246,860; Denominator: Rows 11 – 18 = 6,262,490.

¹⁴ From Table 4: Numerator: Rows 4 – 14, 13, 15, and 17 = 86,570; Denominator: Rows 14, 16, and 18 = 6,275,530.

V. Appendices

Appendix I: Project Phases

In 2004, Dr. Michael Davern from the State Health Access Data Assistance Center (SHADAC), University of Minnesota was funded by the Robert Wood Johnson Foundation to match enrollment data from seven states to the Current Population Survey (CPS) to study why the large discrepancy exists between CPS estimates and Medicaid enrollment counts. During the summer of 2004, Dr. Davern approached Census Bureau and Center for Medicare and Medicaid Services (CMS) officials about participating in the study. Realizing the amount of time that it would take to acquire the state data, Census Bureau officials offered the idea of conducting a national match first using a file they had already acquired: Medicaid Statistical Information System (MSIS) file. Census Bureau officials further suggested that a precursor to the national study could be a simple quality check on both the MSIS and the Medicare Enrollment Database (MEDB) files, which was in scope of current agreements and systems of records notices with CMS.

The federal Health and Human Services Department (HHS) Office of the Assistance Secretary for Planning and Evaluation (ASPE) and SHADAC provided additional funding for the process. This money allowed for the national match and provided an opportunity to analyze the National Health Interview Survey (NHIS) and therefore bring its sponsor, the National Center for Health Statistics (NCHS), in to the study. Note that since NHIS asks a point-in-time question about Medicaid enrollment versus the CPS's "Have you been on Medicaid anytime in the calendar year?," the impact of timing (both for the reference period and its relationship to the moment of survey fielding) can be examined by a comparison of results. Hence, the SNACC Team emerged, which is an acronym for the first initial of each of the participating agencies: SHADAC, NCHS, ASPE, CMS, and the Census Bureau. After the first face-to-face meeting on the project, the SNACC team agreed to break the study into four distinct, but related, phases:

Phase I: Merging the National Level CMS Databases

In Phase I, we create a national database of health-insurance enrollment and evaluate the quality of the information it contains. We create the database by merging the CMS Medicaid Statistical Information System (MSIS) file with the CMS Medicare Enrollment Database (MEDB) file (see Appendix III for an explanation of the research and supporting files). We evaluate the quality of the database by assessing our ability to accurately merge the input files and by comparing the characteristics of the individuals in the database to expectations based on Medicaid eligibility rules and characteristics of the U.S. population. (For more information, please see *Phase I Research Results: Overview of Medicare and Medicaid Files*, February 2007 <http://www.census.gov/did/www/shadac/shadac.html>.)

Phase II: Matching the MSIS to the Current Population Survey (CPS)

In Phase II, we match data from MSIS and the Current Population Survey (CPS) according to a unique person-identifier (the Protected Identity Key or PIK, which replaces Social Security Number or SSN to protect reportee privacy). We supplement the matched records with information from the Medicaid Analytic eXtract (MAX), the Person Characteristic File (PCF), and the Master Address File Auxiliary Reference File (MAF-ARF) and examine why there are

discrepancies between MSIS records of enrollment and CPS reports of Medicaid coverage (See hypotheses outlined below).

Phase III: Matching the State Frame and Person MSIS data to the CPS

In Phase III, we use data from the seven acquired state Medicaid file extracts, CPS, MAX, and the Census Bureau Master Address File (MAF) to enhance the comparison of MSIS enrollment data to CPS health insurance data developed in Phase II. The Census Bureau worked with CMS to negotiate the acquisition of the state Medicaid files that are intended to enhance the capability of finding matches with CPS person records because they include person names and addresses (in contrast with MSIS, which does not include these data elements). We analyze how improved MSIS record person-identification and determination of out-of-CPS-scope status can impact Phase II results.

Phase IV: Matching the MSIS to the National Health Interview Survey (NHIS)

In Phase IV, we repeat the Phase II process, replacing the CPS data with NHIS data. In addition to providing explanations for discrepancies between the national NHIS and MSIS, comparisons to Phase II results will allow the examination of how survey design and implementation affect the quality of the resulting survey data and the estimates derived from them.

Hypotheses about the Discrepancy between Survey Estimates and Administrative Counts

In preparation for Phase II, III, and IV analyses, the team prepared a list of twelve hypotheses explaining the Medicaid enrollment count-survey estimate discrepancies and outlined some possible approaches to evaluating them:

1. Individuals enrolled in Medicaid who were residing in institutional group quarters were not designed to be included in the survey universe. However, they are included in the Medicaid administrative data. Unfortunately, it is not possible to identify all Medicaid enrollees living in group quarters by using the Medicaid administrative data so we cannot directly account for this universe difference. The best approach to evaluating this hypothesis would be to use MAX to identify persons who received Long Term Care (LTC) services (since they mostly receive such care in group facilities) and use the count as a measure of the magnitude of this factor.
2. Individuals enrolled in Medicaid who do not have a usual residence were not designed to be included in the survey universe. Unfortunately, it is not possible to determine the usualness of an enrollee's residence from their Medicaid administrative data, so we cannot directly account for this universe difference. However, these individuals often receive correspondence at local social services offices, soup kitchens or other places so one approach to evaluating this hypothesis would be to run a frequency distribution to determine if many unrelated individuals are attributed to specific addresses.

3. Some individuals are enrolled in Medicaid for a very short period of time (catastrophic health event and related costs made them eligible – or they were resident aliens), so respondents may not consider this coverage. Unfortunately, there is some conceptual question whether these individuals should be considered as “covered” by health insurance. One approach to evaluating the role of this conceptual confusion in the discrepancy between reports and actual receipt of benefits would be to evaluate survey reports after classifying individuals by length of time enrolled.
4. Some groups of Medicaid enrollees have restrictions on the Medicaid benefits they receive regardless of length of enrollment (for example, some dual Medicaid and Medicare enrollees who do not receive the full scope of Medicaid benefits, non-citizens who receive only emergency services, etc.), so respondents for enrollees in these restricted groups may not consider this coverage. However, as in #3 above, there is a conceptual question whether these individuals should be considered as “covered” by health insurance. One approach to evaluating this hypothesis would be to examine reports after classifying enrollees by type of Medicaid benefit they receive.
5. Some Medicaid enrollees are in pre-paid plans (e.g. HMOs - Health Maintenance Organizations, PHPs - Prepaid Health Plans, etc.), so respondents reporting about such enrollees may incorrectly report no coverage because they do not realize that Medicaid is the health insurance that pays for the enrollee’s services. One approach would be to examine reports after classifying individuals by their type of Medicaid plan.
6. Medicaid is a state-based delivery system, so enrollees may be counted more than once if they are enrolled in multiple states or multiple counties in the same state (e.g., because of moving during the year). There are also documented instances where states have changed identification numbers for selected individuals or all state enrollees. While these situations could cause double (or more) counting in administrative data systems, MAX processing attempts to minimize multiple counting within an individual state. However, because MAX and its source, MSIS, are organized into state-specific data sets, some enrollees and beneficiaries may be duplicated across states. CMS began an examination of the extent of multiple counting across states in 2006, so one approach would be to evaluate what CMS results imply about the discrepancy between CPS estimates and Medicaid enrollment counts.
7. Some Medicaid enrollees did not receive Medicaid-provided medical services during the survey reference period, so respondents reporting about such enrollees may incorrectly report no coverage because they may not realize or believe that Medicaid covered the enrollee. One approach to evaluating this hypothesis would be to classify individuals by multiple categories (not eligible, eligible and not enrolled, enrolled and not receiving benefits, enrolled and receiving benefits). However, it is not possible to identify persons eligible and not enrolled in the Medicaid administrative data (because the system captures only those individuals who are actually enrolled). Under this approach it is necessary to use other sources (e.g., the Urban Institute Trim3 model and Census Bureau social

service eligibility modeling work¹) to estimate persons eligible and not enrolled. This is outside the scope of this research but could be addressed in subsequent studies.

8. Medicaid eligibility is not assigned to an entire household (it is case-based or individual-based), so Census Bureau survey households containing multiple families or non-family members are most at risk for incorrect proxy responses. Even when a family and household have a one-to-one correspondence, if the survey respondent is not the applicant for Medicaid, there may be errors. One approach to evaluating this hypothesis is to examine reports after classifying individuals by relationship types.
9. Some Medicaid enrollees have co-insurance or personal or family liability for some share of the cost of their health care, so respondents reporting about them may incorrectly report no coverage because they did not realize or believe that the enrollee was nonetheless covered by Medicaid. One approach would be to review Medicaid data elements that identify individuals with a Medicare deductible and/or coinsurance and that identify individuals with other insurance (this is known in Medicaid as third-party liability).
10. Some respondents may not distinguish the difference between types of health insurance (because they never knew the difference or in the survey they fail to apply the cognitive effort needed to distinguish the difference), so they may, for example, mistake Medicare or stand-alone State Children's Health Insurance Program (SCHIP) for Medicaid and vice-versa. One approach to evaluating this hypothesis would be to check survey results after classifying enrollees by dual eligible status and stand-alone SCHIP enrollment status.
11. Respondents may systematically make recall errors reporting about individuals who did not receive medical services from Medicaid or were not enrolled in the month in which they were surveyed.² One approach would be to check survey results after classifying enrollees by their Medicaid experience in the month of the survey.
12. Survey procedures may present biases for poorer populations, including the coverage of the survey frame (derived from the Master Address File or MAF) or non-response weighting. One approach to checking the MAF would be to match all Medicaid address records to the MAF to determine what proportion of Medicaid cases were out of the survey universe and study how those cases may differ from those that were correctly included in the survey estimate.

High-Level Research Requirements

The previous twelve hypotheses support the following high-level research requirements:

¹ See <http://www.fcsn.gov/07papers/Resnick.VII-A.pdf> for additional information on eligibility modeling.

² Mah and Resnick (2007) analyzed possible response biases in CPS between Medicaid enrollees and non-enrollees. Their research, presented at the 2007 Joint Statistical Meetings, found that any such bias is relatively small.

- Identify the size of the discrepancy between MSIS enrollment counts and CPS enrollment estimates.
- Provide information and demographics (including family income as a percentage of the poverty threshold) on the false negatives (those who are enrolled in Medicaid but respond “no” on the CPS).
- Provide information and demographics (including family income as a percentage of the poverty threshold) on the unconfirmed persons reported enrolled (those who are reported enrolled in Medicaid on CPS but for whom confirmation by linkage to an MSIS record does not exist). Note that only some of these persons are false positives. Others may be unconfirmable due to incomplete linking data on MSIS.
- Provide information on factors that are contributing to both sources of mis-alignment.
- Provide some measure of “time since covered” or “intensity of coverage.”
- Provide information on what types of coverage false negatives are receiving, since they are not reporting Medicaid coverage.

Summarized Results from Previous Phases

Phase I: Merging the National Level CMS Databases

These are the most relevant conclusions from the exploratory analysis of CMS databases, MSIS (Medicaid) and MEDB (Medicare):

- MSIS person-identification validation varied significantly by state.
- About 2.2 percent of the MSIS uniquely-identified individuals have records in multiple states simultaneously, suggesting probable duplication.
- The demographic and programmatic characteristics of individuals in the database of health-insurance enrollment correspond to the characteristics of the broader U.S. population and eligibility rules for Medicaid and Medicare.
- MSIS file is of high enough quality to have confidence in our subsequent research into the discrepancy between Medicaid enrollees and reported enrollment in the CPS.

Phase II: Matching the MSIS to the Current Population Survey (CPS)

In Phase II we saw that after steps had been taken to bring the MSIS universe and CPS universe into alignment, the largest source of the discrepancy between the MSIS count of identified Year 2000 enrollees of 36.2 million and the CPS/ASEC estimate of 26.1 Year 2000 enrollees was the misclassification by the CPS/ASEC of the enrollment status of MSIS-reported enrollees as non-enrollees. Most of this misclassification is due to respondent misreports of the Medicaid enrollment status of their household members. CPS/ASEC imputations of Medicaid enrollment status largely mirrored that of explicit respondents in terms of aggregate accuracy. Had all MSIS-recorded full benefit MSIS enrollees been classified as such on CPS/ASEC, the estimate of Year 2000 enrollees would have been 33.4 million.

Descriptive statistics and regression analysis shows that mis-reporting tends to align with the non-salience of Medicaid enrollment status to the household respondent. Medicaid enrollment

reporting was more accurate for persons enrolled at the time of the interview, for persons with coverage during a large portion of reference period, and for persons covered near the time of interview. Also, if the referent received medical services, reporting was more accurate than otherwise. It remains unclear why Medicaid enrollees are frequently not identified as such on CPS/ASEC, but it is apparent that the more factors available to remind the respondent of this coverage, the less likely they were to not report it (that is make a false-negative report).

Appendix II: Phase II CPS/ASEC Underestimate Discrepancy Accounting Table (Showing Effect of Phase III Adjustment)

Pages 1-8: Prior to Phase III Adjustment

Pages 9-16: Subsequent to Phase III Adjustment

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight**

State=U.S. Total

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Total	45,050,000	43,650,000	39,750,000	39,600,000	38,150,000	36,200,000	29,550,000	26,050,000	18,600,000
Age 0 - 5	9,590,000	9,340,000	9,290,000	9,290,000	8,840,000	7,820,000	5,860,000	5,100,000	4,020,000
Age 6 - 14	10,450,000	9,790,000	9,640,000	9,640,000	9,240,000	8,990,000	7,200,000	6,000,000	4,460,000
Age 15 - 17	2,720,000	2,480,000	2,310,000	2,310,000	2,230,000	2,150,000	1,740,000	1,480,000	1,040,000
Age 18 - 44	13,750,000	13,550,000	10,950,000	10,950,000	10,600,000	10,250,000	7,040,000	6,500,000	4,420,000
Age 45 - 64	4,020,000	4,010,000	3,650,000	3,630,000	3,550,000	3,500,000	3,620,000	3,280,000	2,200,000
Age 65+	4,380,000	4,380,000	3,770,000	3,640,000	3,580,000	3,470,000	3,180,000	2,880,000	2,220,000
Age N/A	139,000	135,000	129,000	129,000	129,000	21,400	900,000	800,000	260,000
White	27,600,000	26,750,000	25,300,000	25,150,000	24,250,000	23,650,000	19,900,000	17,300,000	12,450,000
Black	12,000,000	11,600,000	11,150,000	11,100,000	10,700,000	10,250,000	7,740,000	7,120,000	5,020,000
AIAN	806,000	781,000	756,000	755,000	709,000	680,000	620,000	580,000	460,000
API	1,910,000	1,880,000	1,740,000	1,740,000	1,700,000	1,640,000	1,280,000	1,080,000	700,000
Race Unknown	2,700,000	2,680,000	795,000	795,000	795,000	0	-	-	-
Male	18,550,000	17,850,000	16,950,000	16,900,000	16,250,000	15,450,000	13,050,000	11,400,000	7,940,000
Female	26,500,000	25,800,000	22,800,000	22,700,000	21,900,000	20,750,000	16,450,000	14,700,000	10,650,000
Hispanic	10,500,000	10,350,000	8,260,000	8,250,000	7,980,000	7,490,000	6,560,000	5,780,000	4,260,000
Non-Hispanic	32,300,000	31,150,000	30,000,000	29,850,000	28,750,000	28,750,000	23,000,000	20,300,000	14,350,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight**

State=State Summary

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Alabama	686,000	675,000	624,000	623,000	620,000	600,000	520,000	460,000	340,000
Alaska	106,000	95,100	95,000	94,900	94,800	89,900	80,000	60,000	40,000
Arizona	646,000	646,000	638,000	637,000	635,000	586,000	540,000	480,000	360,000
Arkansas	474,000	473,000	452,000	451,000	451,000	382,000	320,000	280,000	200,000
California	8,080,000	8,050,000	5,630,000	5,620,000	5,580,000	5,320,000	4,640,000	4,260,000	3,020,000
Colorado	347,000	347,000	317,000	315,000	315,000	292,000	260,000	200,000	140,000
Connecticut	396,000	304,000	299,000	294,000	293,000	284,000	280,000	200,000	140,000
Delaware	117,000	117,000	109,000	109,000	109,000	103,000	60,000	60,000	40,000
District of Columbia	137,000	134,000	133,000	133,000	133,000	127,000	80,000	80,000	60,000
Florida	2,120,000	2,080,000	1,920,000	1,910,000	1,910,000	1,850,000	1,680,000	1,340,000	920,000
Georgia	1,410,000	1,200,000	1,160,000	1,160,000	1,100,000	992,000	820,000	740,000	460,000
Hawaii	174,000	173,000	173,000	173,000	173,000	170,000	100,000	60,000	40,000
Idaho	141,000	132,000	130,000	130,000	129,000	124,000	120,000	100,000	80,000
Illinois	1,660,000	1,610,000	1,570,000	1,560,000	1,550,000	1,470,000	1,040,000	880,000	660,000
Indiana	711,000	681,000	657,000	651,000	651,000	637,000	340,000	320,000	260,000
Iowa	290,000	285,000	277,000	275,000	274,000	269,000	180,000	140,000	100,000
Kansas	240,000	240,000	233,000	232,000	232,000	226,000	180,000	140,000	100,000
Kentucky	715,000	671,000	639,000	638,000	612,000	593,000	420,000	360,000	260,000
Louisiana	787,000	750,000	713,000	710,000	710,000	659,000	540,000	500,000	320,000
Maine	202,000	194,000	189,000	188,000	188,000	184,000	140,000	140,000	100,000
Maryland	634,000	553,000	500,000	498,000	498,000	464,000	320,000	240,000	180,000
Massachusetts	1,060,000	996,000	872,000	864,000	863,000	823,000	860,000	820,000	600,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight**

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Michigan	1,290,000	1,270,000	1,250,000	1,240,000	1,240,000	1,190,000	960,000	920,000	620,000
Minnesota	574,000	574,000	554,000	552,000	540,000	528,000	360,000	340,000	220,000
Mississippi	565,000	426,000	423,000	422,000	422,000	395,000	420,000	380,000	300,000
Missouri	925,000	871,000	859,000	855,000	855,000	816,000	520,000	480,000	420,000
Montana	97,500	89,000	76,400	76,200	76,200	48,500	100,000	80,000	60,000
Nebraska	219,000	212,000	211,000	210,000	210,000	204,000	120,000	120,000	100,000
Nevada	132,000	132,000	124,000	123,000	123,000	115,000	140,000	100,000	60,000
New Hampshire	98,900	98,800	98,000	97,200	97,200	94,600	80,000	80,000	60,000
New Jersey	939,000	858,000	811,000	804,000	771,000	678,000	620,000	540,000	320,000
New Mexico	382,000	378,000	353,000	352,000	352,000	337,000	260,000	240,000	160,000
New York	3,260,000	3,260,000	3,210,000	3,190,000	3,150,000	2,770,000	2,720,000	2,360,000	1,680,000
North Carolina	1,230,000	1,160,000	1,100,000	1,090,000	1,090,000	1,070,000	820,000	740,000	580,000
North Dakota	59,100	56,600	55,600	55,400	54,600	53,900	40,000	40,000	20,000
Ohio	1,360,000	1,310,000	1,290,000	1,280,000	1,260,000	1,230,000	900,000	860,000	660,000
Oklahoma	658,000	617,000	597,000	596,000	479,000	450,000	320,000	280,000	220,000
Oregon	534,000	518,000	485,000	484,000	482,000	463,000	380,000	340,000	240,000
Pennsylvania	1,670,000	1,670,000	1,580,000	1,570,000	1,570,000	1,540,000	1,120,000	820,000	500,000
Rhode Island	172,000	164,000	158,000	157,000	157,000	154,000	140,000	120,000	100,000
South Carolina	754,000	713,000	627,000	627,000	626,000	603,000	400,000	360,000	300,000
South Dakota	92,700	89,200	85,700	85,300	84,000	81,300	40,000	40,000	40,000
Tennessee	1,430,000	1,420,000	1,390,000	1,390,000	1,380,000	1,360,000	1,020,000	1,000,000	760,000
Texas	2,620,000	2,590,000	2,410,000	2,400,000	2,400,000	2,280,000	1,840,000	1,580,000	1,100,000
Utah	209,000	189,000	184,000	184,000	184,000	176,000	180,000	160,000	120,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight**

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Vermont	142,000	140,000	128,000	127,000	127,000	126,000	100,000	100,000	80,000
Virginia	684,000	652,000	611,000	608,000	580,000	539,000	460,000	420,000	300,000
Washington	908,000	908,000	902,000	901,000	900,000	861,000	660,000	520,000	340,000
West Virginia	326,000	325,000	315,000	314,000	314,000	293,000	220,000	200,000	140,000
Wisconsin	583,000	563,000	557,000	553,000	551,000	533,000	440,000	400,000	300,000
Wyoming	46,300	46,300	44,300	44,000	44,000	41,800	40,000	40,000	20,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight**

State=California

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Total	8,080,000	8,050,000	5,630,000	5,620,000	5,580,000	5,320,000	4,640,000	4,260,000	3,020,000
Age 0 - 5	1,250,000	1,250,000	1,220,000	1,220,000	1,170,000	1,050,000	840,000	760,000	580,000
Age 6 - 14	1,480,000	1,470,000	1,360,000	1,360,000	1,360,000	1,320,000	1,160,000	1,040,000	760,000
Age 15 - 17	477,000	472,000	328,000	328,000	327,000	311,000	300,000	280,000	200,000
Age 18 - 44	3,530,000	3,530,000	1,530,000	1,530,000	1,530,000	1,480,000	1,140,000	1,080,000	740,000
Age 45 - 64	648,000	648,000	548,000	546,000	546,000	539,000	560,000	520,000	340,000
Age 65+	684,000	684,000	651,000	643,000	643,000	630,000	460,000	460,000	340,000
Age N/A	5,400	5,400	150	150	150	50	180,000	160,000	60,000
White	4,290,000	4,270,000	3,820,000	3,810,000	3,770,000	3,720,000	3,540,000	3,220,000	2,300,000
Black	941,000	940,000	815,000	814,000	808,000	779,000	460,000	440,000	300,000
AIAN	108,000	107,000	96,500	96,400	95,500	94,800	100,000	100,000	80,000
API	847,000	844,000	741,000	740,000	737,000	725,000	540,000	500,000	360,000
Race Unknown	1,900,000	1,890,000	162,000	162,000	162,000	0	-	-	-
Male	2,930,000	2,920,000	2,520,000	2,510,000	2,490,000	2,380,000	2,040,000	1,840,000	1,260,000
Female	5,150,000	5,140,000	3,120,000	3,110,000	3,090,000	2,940,000	2,600,000	2,420,000	1,760,000
Hispanic	4,390,000	4,370,000	2,600,000	2,600,000	2,570,000	2,440,000	2,280,000	2,080,000	1,500,000
Non-Hispanic	2,940,000	2,930,000	2,910,000	2,900,000	2,880,000	2,880,000	2,360,000	2,180,000	1,540,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight**

State=Florida

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Total	2,120,000	2,080,000	1,920,000	1,910,000	1,910,000	1,850,000	1,680,000	1,340,000	920,000
Age 0 - 5	479,000	479,000	475,000	475,000	475,000	450,000	280,000	200,000	160,000
Age 6 - 14	490,000	489,000	485,000	485,000	485,000	481,000	400,000	260,000	180,000
Age 15 - 17	123,000	115,000	113,000	113,000	113,000	111,000	100,000	80,000	60,000
Age 18 - 44	583,000	560,000	469,000	468,000	468,000	451,000	360,000	320,000	200,000
Age 45 - 64	176,000	173,000	149,000	148,000	147,000	146,000	180,000	160,000	100,000
Age 65+	267,000	267,000	225,000	218,000	218,000	213,000	320,000	260,000	220,000
Age N/A	350	300	300	300	300	100	60,000	60,000	20,000
White	1,300,000	1,270,000	1,150,000	1,150,000	1,150,000	1,130,000	1,160,000	900,000	620,000
Black	745,000	736,000	698,000	696,000	695,000	681,000	480,000	440,000	280,000
AIAN	10,300	10,200	9,500	9,500	9,500	9,450	20,000	20,000	0
API	32,600	32,100	29,900	29,800	29,800	29,400	20,000	0	0
Race Unknown	30,900	29,400	24,800	24,800	24,800	0	-	-	-
Male	891,000	871,000	821,000	819,000	818,000	800,000	760,000	600,000	400,000
Female	1,230,000	1,210,000	1,090,000	1,090,000	1,090,000	1,050,000	920,000	740,000	520,000
Hispanic	548,000	531,000	493,000	492,000	492,000	472,000	480,000	400,000	300,000
Non-Hispanic	1,530,000	1,510,000	1,390,000	1,380,000	1,380,000	1,380,000	1,200,000	960,000	620,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight**

State=New York

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Total	3,260,000	3,260,000	3,210,000	3,190,000	3,150,000	2,770,000	2,720,000	2,360,000	1,680,000
Age 0 - 5	559,000	559,000	558,000	558,000	547,000	438,000	420,000	340,000	240,000
Age 6 - 14	615,000	615,000	613,000	613,000	602,000	566,000	660,000	500,000	380,000
Age 15 - 17	158,000	156,000	155,000	155,000	151,000	141,000	140,000	100,000	80,000
Age 18 - 44	992,000	989,000	976,000	975,000	962,000	883,000	660,000	620,000	440,000
Age 45 - 64	422,000	422,000	412,000	410,000	405,000	392,000	380,000	360,000	260,000
Age 65+	398,000	398,000	380,000	361,000	357,000	334,000	360,000	320,000	260,000
Age N/A	120,000	120,000	120,000	120,000	120,000	16,100	100,000	80,000	20,000
White	1,770,000	1,770,000	1,740,000	1,720,000	1,700,000	1,630,000	1,560,000	1,320,000	880,000
Black	1,040,000	1,040,000	1,030,000	1,030,000	1,010,000	954,000	960,000	900,000	700,000
AIAN	27,900	27,800	27,200	27,100	26,800	22,800	40,000	40,000	20,000
API	182,000	182,000	180,000	179,000	176,000	160,000	160,000	120,000	60,000
Race Unknown	240,000	240,000	236,000	236,000	236,000	0	-	-	-
Male	1,370,000	1,360,000	1,350,000	1,340,000	1,320,000	1,210,000	1,160,000	1,000,000	700,000
Female	1,900,000	1,890,000	1,870,000	1,850,000	1,830,000	1,560,000	1,560,000	1,380,000	980,000
Hispanic	910,000	909,000	903,000	901,000	886,000	820,000	840,000	760,000	560,000
Non-Hispanic	2,040,000	2,030,000	2,000,000	1,980,000	1,950,000	1,950,000	1,880,000	1,600,000	1,120,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight**

State=Texas

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Total	2,620,000	2,590,000	2,410,000	2,400,000	2,400,000	2,280,000	1,840,000	1,580,000	1,100,000
Age 0 - 5	779,000	779,000	777,000	777,000	776,000	675,000	480,000	420,000	360,000
Age 6 - 14	618,000	617,000	615,000	615,000	615,000	608,000	440,000	340,000	260,000
Age 15 - 17	150,000	140,000	137,000	137,000	137,000	135,000	140,000	120,000	60,000
Age 18 - 44	577,000	559,000	494,000	493,000	493,000	489,000	360,000	320,000	180,000
Age 45 - 64	177,000	177,000	151,000	150,000	149,000	149,000	220,000	200,000	100,000
Age 65+	321,000	321,000	235,000	229,000	229,000	224,000	180,000	160,000	120,000
Age N/A	50	50	0	0	0	0	60,000	40,000	20,000
White	1,900,000	1,880,000	1,770,000	1,760,000	1,760,000	1,730,000	1,460,000	1,240,000	900,000
Black	534,000	529,000	502,000	501,000	500,000	483,000	300,000	260,000	180,000
AIAN	27,800	27,600	26,200	26,200	26,200	25,500	20,000	20,000	20,000
API	48,600	48,100	46,300	46,200	46,200	44,300	60,000	60,000	0
Race Unknown	111,000	111,000	64,200	64,200	64,200	0	-	-	-
Male	1,100,000	1,080,000	1,030,000	1,030,000	1,030,000	969,000	860,000	720,000	480,000
Female	1,520,000	1,510,000	1,380,000	1,370,000	1,370,000	1,310,000	980,000	860,000	600,000
Hispanic	1,390,000	1,370,000	1,270,000	1,270,000	1,270,000	1,200,000	1,040,000	880,000	660,000
Non-Hispanic	1,170,000	1,160,000	1,080,000	1,080,000	1,080,000	1,080,000	800,000	700,000	420,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight, Phase III Analysis Modified**

State=U.S. Total

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Total	45,050,000	43,650,000	39,750,000	38,650,000	37,250,000	35,300,000	29,550,000	26,050,000	18,600,000
Age 0 - 5	9,590,000	9,340,000	9,290,000	9,280,000	8,840,000	7,810,000	5,860,000	5,100,000	4,020,000
Age 6 - 14	10,450,000	9,790,000	9,640,000	9,630,000	9,230,000	8,970,000	7,200,000	6,000,000	4,460,000
Age 15 - 17	2,720,000	2,480,000	2,310,000	2,280,000	2,200,000	2,120,000	1,740,000	1,480,000	1,040,000
Age 18 - 44	13,750,000	13,550,000	10,950,000	10,800,000	10,450,000	10,100,000	7,040,000	6,500,000	4,420,000
Age 45 - 64	4,020,000	4,010,000	3,650,000	3,510,000	3,430,000	3,380,000	3,620,000	3,280,000	2,200,000
Age 65+	4,380,000	4,380,000	3,770,000	3,030,000	2,980,000	2,870,000	3,180,000	2,880,000	2,220,000
Age N/A	139,000	135,000	129,000	129,000	129,000	21,400	900,000	800,000	260,000
White	27,600,000	26,750,000	25,300,000	24,450,000	23,550,000	22,900,000	19,900,000	17,300,000	12,450,000
Black	12,000,000	11,600,000	11,150,000	10,950,000	10,550,000	10,100,000	7,740,000	7,120,000	5,020,000
AIAN	806,000	781,000	756,000	746,000	701,000	672,000	620,000	580,000	460,000
API	1,910,000	1,880,000	1,740,000	1,730,000	1,680,000	1,630,000	1,280,000	1,080,000	700,000
Race Unknown	2,700,000	2,680,000	795,000	795,000	795,000	0	-	-	-
Male	18,550,000	17,850,000	16,950,000	16,550,000	15,900,000	15,100,000	13,050,000	11,400,000	7,940,000
Female	26,500,000	25,800,000	22,800,000	22,100,000	21,350,000	20,150,000	16,450,000	14,700,000	10,650,000
Hispanic	10,500,000	10,350,000	8,260,000	8,190,000	7,920,000	7,430,000	6,560,000	5,780,000	4,260,000
Non-Hispanic	32,300,000	31,150,000	30,000,000	29,000,000	27,850,000	27,850,000	23,000,000	20,300,000	14,350,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight, Phase III Analysis Modified**

State=State Summary

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Alabama	686,000	675,000	624,000	607,000	605,000	585,000	520,000	460,000	340,000
Alaska	106,000	95,100	95,000	94,000	94,000	89,100	80,000	60,000	40,000
Arizona	646,000	646,000	638,000	631,000	629,000	580,000	540,000	480,000	360,000
Arkansas	474,000	473,000	452,000	439,000	438,000	370,000	320,000	280,000	200,000
California	8,080,000	8,050,000	5,630,000	5,540,000	5,500,000	5,240,000	4,640,000	4,260,000	3,020,000
Colorado	347,000	347,000	317,000	306,000	306,000	283,000	260,000	200,000	140,000
Connecticut	396,000	304,000	299,000	279,000	278,000	269,000	280,000	200,000	140,000
Delaware	117,000	117,000	109,000	106,000	106,000	100,000	60,000	60,000	40,000
District of Columbia	137,000	134,000	133,000	130,000	130,000	124,000	80,000	80,000	60,000
Florida	2,120,000	2,080,000	1,920,000	1,870,000	1,870,000	1,810,000	1,680,000	1,340,000	920,000
Georgia	1,410,000	1,200,000	1,160,000	1,140,000	1,080,000	970,000	820,000	740,000	460,000
Hawaii	174,000	173,000	173,000	171,000	171,000	168,000	100,000	60,000	40,000
Idaho	141,000	132,000	130,000	126,000	126,000	121,000	120,000	100,000	80,000
Illinois	1,660,000	1,610,000	1,570,000	1,510,000	1,500,000	1,420,000	1,040,000	880,000	660,000
Indiana	711,000	681,000	657,000	628,000	628,000	614,000	340,000	320,000	260,000
Iowa	290,000	285,000	277,000	261,000	260,000	255,000	180,000	140,000	100,000
Kansas	240,000	240,000	233,000	221,000	221,000	215,000	180,000	140,000	100,000
Kentucky	715,000	671,000	639,000	621,000	595,000	576,000	420,000	360,000	260,000
Louisiana	787,000	750,000	713,000	688,000	688,000	637,000	540,000	500,000	320,000
Maine	202,000	194,000	189,000	183,000	183,000	179,000	140,000	140,000	100,000
Maryland	634,000	553,000	500,000	484,000	483,000	450,000	320,000	240,000	180,000
Massachusetts	1,060,000	996,000	872,000	834,000	833,000	793,000	860,000	820,000	600,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight, Phase III Analysis Modified**

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Michigan	1,290,000	1,270,000	1,250,000	1,210,000	1,210,000	1,160,000	960,000	920,000	620,000
Minnesota	574,000	574,000	554,000	530,000	518,000	507,000	360,000	340,000	220,000
Mississippi	565,000	426,000	423,000	410,000	410,000	383,000	420,000	380,000	300,000
Missouri	925,000	871,000	859,000	830,000	830,000	791,000	520,000	480,000	420,000
Montana	97,500	89,000	76,400	72,900	72,900	45,200	100,000	80,000	60,000
Nebraska	219,000	212,000	211,000	202,000	202,000	196,000	120,000	120,000	100,000
Nevada	132,000	132,000	124,000	121,000	121,000	112,000	140,000	100,000	60,000
New Hampshire	98,900	98,800	98,000	93,200	93,200	90,600	80,000	80,000	60,000
New Jersey	939,000	858,000	811,000	782,000	749,000	646,000	620,000	540,000	320,000
New Mexico	382,000	378,000	353,000	348,000	348,000	333,000	260,000	240,000	160,000
New York	3,260,000	3,260,000	3,210,000	3,100,000	3,060,000	2,690,000	2,720,000	2,360,000	1,680,000
North Carolina	1,230,000	1,160,000	1,100,000	1,070,000	1,070,000	1,040,000	820,000	740,000	580,000
North Dakota	59,100	56,600	55,600	52,000	51,200	50,500	40,000	40,000	20,000
Ohio	1,360,000	1,310,000	1,290,000	1,230,000	1,220,000	1,190,000	900,000	860,000	660,000
Oklahoma	658,000	617,000	597,000	579,000	465,000	435,000	320,000	280,000	220,000
Oregon	534,000	518,000	485,000	475,000	473,000	453,000	380,000	340,000	240,000
Pennsylvania	1,670,000	1,670,000	1,580,000	1,530,000	1,530,000	1,500,000	1,120,000	820,000	500,000
Rhode Island	172,000	164,000	158,000	153,000	153,000	150,000	140,000	120,000	100,000
South Carolina	754,000	713,000	627,000	615,000	615,000	591,000	400,000	360,000	300,000
South Dakota	92,700	89,200	85,700	81,100	80,000	77,300	40,000	40,000	40,000
Tennessee	1,430,000	1,420,000	1,390,000	1,370,000	1,370,000	1,340,000	1,020,000	1,000,000	760,000
Texas	2,620,000	2,590,000	2,410,000	2,340,000	2,340,000	2,220,000	1,840,000	1,580,000	1,100,000
Utah	209,000	189,000	184,000	180,000	180,000	173,000	180,000	160,000	120,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight, Phase III Analysis Modified**

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Vermont	142,000	140,000	128,000	125,000	125,000	124,000	100,000	100,000	80,000
Virginia	684,000	652,000	611,000	590,000	562,000	521,000	460,000	420,000	300,000
Washington	908,000	908,000	902,000	888,000	887,000	848,000	660,000	520,000	340,000
West Virginia	326,000	325,000	315,000	308,000	307,000	286,000	220,000	200,000	140,000
Wisconsin	583,000	563,000	557,000	528,000	526,000	509,000	440,000	400,000	300,000
Wyoming	46,300	46,300	44,300	42,500	42,500	40,300	40,000	40,000	20,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight, Phase III Analysis Modified**

State=California

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Total	8,080,000	8,050,000	5,630,000	5,540,000	5,500,000	5,240,000	4,640,000	4,260,000	3,020,000
Age 0 - 5	1,250,000	1,250,000	1,220,000	1,220,000	1,170,000	1,050,000	840,000	760,000	580,000
Age 6 - 14	1,480,000	1,470,000	1,360,000	1,360,000	1,360,000	1,320,000	1,160,000	1,040,000	760,000
Age 15 - 17	477,000	472,000	328,000	325,000	324,000	308,000	300,000	280,000	200,000
Age 18 - 44	3,530,000	3,530,000	1,530,000	1,510,000	1,500,000	1,450,000	1,140,000	1,080,000	740,000
Age 45 - 64	648,000	648,000	548,000	533,000	533,000	526,000	560,000	520,000	340,000
Age 65+	684,000	684,000	651,000	602,000	602,000	589,000	460,000	460,000	340,000
Age N/A	5,400	5,400	150	150	150	50	180,000	160,000	60,000
White	4,290,000	4,270,000	3,820,000	3,750,000	3,710,000	3,660,000	3,540,000	3,220,000	2,300,000
Black	941,000	940,000	815,000	801,000	796,000	766,000	460,000	440,000	300,000
AIAN	108,000	107,000	96,500	95,300	94,400	93,700	100,000	100,000	80,000
API	847,000	844,000	741,000	734,000	731,000	719,000	540,000	500,000	360,000
Race Unknown	1,900,000	1,890,000	162,000	162,000	162,000	0	-	-	-
Male	2,930,000	2,920,000	2,520,000	2,480,000	2,450,000	2,340,000	2,040,000	1,840,000	1,260,000
Female	5,150,000	5,140,000	3,120,000	3,070,000	3,050,000	2,900,000	2,600,000	2,420,000	1,760,000
Hispanic	4,390,000	4,370,000	2,600,000	2,580,000	2,550,000	2,420,000	2,280,000	2,080,000	1,500,000
Non-Hispanic	2,940,000	2,930,000	2,910,000	2,830,000	2,820,000	2,820,000	2,360,000	2,180,000	1,540,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight, Phase III Analysis Modified**

State=Florida

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Total	2,120,000	2,080,000	1,920,000	1,870,000	1,870,000	1,810,000	1,680,000	1,340,000	920,000
Age 0 - 5	479,000	479,000	475,000	475,000	474,000	450,000	280,000	200,000	160,000
Age 6 - 14	490,000	489,000	485,000	485,000	485,000	480,000	400,000	260,000	180,000
Age 15 - 17	123,000	115,000	113,000	112,000	112,000	110,000	100,000	80,000	60,000
Age 18 - 44	583,000	560,000	469,000	462,000	462,000	444,000	360,000	320,000	200,000
Age 45 - 64	176,000	173,000	149,000	143,000	143,000	141,000	180,000	160,000	100,000
Age 65+	267,000	267,000	225,000	192,000	192,000	187,000	320,000	260,000	220,000
Age N/A	350	300	300	300	300	100	60,000	60,000	20,000
White	1,300,000	1,270,000	1,150,000	1,120,000	1,120,000	1,100,000	1,160,000	900,000	620,000
Black	745,000	736,000	698,000	688,000	687,000	672,000	480,000	440,000	280,000
AIAN	10,300	10,200	9,500	9,400	9,400	9,350	20,000	20,000	0
API	32,600	32,100	29,900	29,500	29,500	29,100	20,000	0	0
Race Unknown	30,900	29,400	24,800	24,800	24,800	0	-	-	-
Male	891,000	871,000	821,000	805,000	804,000	786,000	760,000	600,000	400,000
Female	1,230,000	1,210,000	1,090,000	1,060,000	1,060,000	1,030,000	920,000	740,000	520,000
Hispanic	548,000	531,000	493,000	488,000	488,000	468,000	480,000	400,000	300,000
Non-Hispanic	1,530,000	1,510,000	1,390,000	1,350,000	1,350,000	1,350,000	1,200,000	960,000	620,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight, Phase III Analysis Modified**

State=New York

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Total	3,260,000	3,260,000	3,210,000	3,100,000	3,060,000	2,690,000	2,720,000	2,360,000	1,680,000
Age 0 - 5	559,000	559,000	558,000	558,000	547,000	438,000	420,000	340,000	240,000
Age 6 - 14	615,000	615,000	613,000	611,000	601,000	565,000	660,000	500,000	380,000
Age 15 - 17	158,000	156,000	155,000	152,000	149,000	138,000	140,000	100,000	80,000
Age 18 - 44	992,000	989,000	976,000	956,000	944,000	865,000	660,000	620,000	440,000
Age 45 - 64	422,000	422,000	412,000	398,000	394,000	381,000	380,000	360,000	260,000
Age 65+	398,000	398,000	380,000	310,000	307,000	283,000	360,000	320,000	260,000
Age N/A	120,000	120,000	120,000	120,000	120,000	16,100	100,000	80,000	20,000
White	1,770,000	1,770,000	1,740,000	1,660,000	1,630,000	1,570,000	1,560,000	1,320,000	880,000
Black	1,040,000	1,040,000	1,030,000	1,010,000	989,000	933,000	960,000	900,000	700,000
AIAN	27,900	27,800	27,200	26,800	26,500	22,500	40,000	40,000	20,000
API	182,000	182,000	180,000	177,000	174,000	158,000	160,000	120,000	60,000
Race Unknown	240,000	240,000	236,000	236,000	236,000	0	-	-	-
Male	1,370,000	1,360,000	1,350,000	1,300,000	1,280,000	1,170,000	1,160,000	1,000,000	700,000
Female	1,900,000	1,890,000	1,870,000	1,800,000	1,780,000	1,510,000	1,560,000	1,380,000	980,000
Hispanic	910,000	909,000	903,000	892,000	877,000	811,000	840,000	760,000	560,000
Non-Hispanic	2,040,000	2,030,000	2,000,000	1,900,000	1,870,000	1,870,000	1,880,000	1,600,000	1,120,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

**Phase II, Table 2: Medicaid Population Size, Comparison of MSIS Counts to CPS Estimates
CY 2000 Expanded Sample, Original Weight, Phase III Analysis Modified**

State=Texas

Selected Characteristics	MSIS Total A	MSIS Total B	MSIS Total C	MSIS Total D	MSIS Total E	MSIS Total F	CPS Total A	CPS Total B	CPS Total C
Total	2,620,000	2,590,000	2,410,000	2,340,000	2,340,000	2,220,000	1,840,000	1,580,000	1,100,000
Age 0 - 5	779,000	779,000	777,000	777,000	776,000	675,000	480,000	420,000	360,000
Age 6 - 14	618,000	617,000	615,000	614,000	614,000	607,000	440,000	340,000	260,000
Age 15 - 17	150,000	140,000	137,000	136,000	135,000	134,000	140,000	120,000	60,000
Age 18 - 44	577,000	559,000	494,000	486,000	485,000	482,000	360,000	320,000	180,000
Age 45 - 64	177,000	177,000	151,000	142,000	142,000	141,000	220,000	200,000	100,000
Age 65+	321,000	321,000	235,000	188,000	188,000	183,000	180,000	160,000	120,000
Age N/A	50	50	0	0	0	0	60,000	40,000	20,000
White	1,900,000	1,880,000	1,770,000	1,720,000	1,710,000	1,680,000	1,460,000	1,240,000	900,000
Black	534,000	529,000	502,000	491,000	490,000	473,000	300,000	260,000	180,000
AIAN	27,800	27,600	26,200	25,800	25,700	25,000	20,000	20,000	20,000
API	48,600	48,100	46,300	45,700	45,700	43,800	60,000	60,000	0
Race Unknown	111,000	111,000	64,200	64,200	64,200	0	-	-	-
Male	1,100,000	1,080,000	1,030,000	1,010,000	1,010,000	948,000	860,000	720,000	480,000
Female	1,520,000	1,510,000	1,380,000	1,330,000	1,330,000	1,270,000	980,000	860,000	600,000
Hispanic	1,390,000	1,370,000	1,270,000	1,260,000	1,260,000	1,190,000	1,040,000	880,000	660,000
Non-Hispanic	1,170,000	1,160,000	1,080,000	1,030,000	1,030,000	1,030,000	800,000	700,000	420,000

1. All CPS Estimates are of weighted totals.
2. MSIS Total A is total enrolled in MSIS.
3. MSIS Total B is total MSIS without SCHIP.
4. MSIS Total C is total enrolled for full benefits.
5. MSIS Total D is total enrolled for full benefits, excluding residents of institutional-GQs.
6. MSIS Total E is total enrolled for full benefits, excluding residents of institutional-GQs and duplicative client accounts.
7. MSIS Total F is Total E excluding un-identified clients: those having records with un-validated SSNs.
8. CPS Total A uses MCAID.
9. CPS Total B uses Augmented-CAID.
10. CPS Total C uses explicitly reported in Medicaid: Augmented-CAID (excluding imputed or edited responses).

Hyphen (-) in cell indicates that the category (for the row) did not exist in that source of data.

Due to rounding, total values may not equal column or row sums.

Appendix III: Description of Files and Methods

Research Files

Medicaid Statistical Information System (MSIS)

The Medicaid Statistical Information System (MSIS) is the basic source of state-submitted eligibility and claims data on the Medicaid population, their characteristics, utilization, and payments. Beginning with Fiscal Year 1999, the Balanced Budget Act (BBA) of 1997 requires states to submit all their eligibility and claims data to CMS on a quarterly basis through the MSIS.

Medicaid Analytic eXtract File (MAX)

The MAX – formerly known as State Medicaid Research Files (SMRFs) – are a set of person-level data files on Medicaid eligibility, service utilization, and payments. The MAX data are extracted from the MSIS. The MAX development process combines MSIS initial claims, interim claims, voids, and adjustments for a given service into this final action event. Unlike fiscal-based MSIS quarterly files, MAX is an annual calendar year file.

Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC)

The CPS ASEC is an annual survey of approximately 78,000 households nationwide. The ASEC, formerly known as the March Supplement to the Current Population Survey is a basic source for health insurance coverage data. The population represented is the civilian non-institutional population living in the United States. Most of the data from the CPS ASEC were collected in March (with some data collected in February and April).

In September 2000, the CPS began an expansion of its sample to accommodate increased funding provided by the State Children's Health Insurance Program (SCHIP). Although focused primarily on the ASEC sample, the SCHIP funding provided for an approximately 20 percent increase in the basic CPS sample. This expansion, which occurred in 31 small states and the District of Columbia, was completed in November 2000. Over the next several months, both the Bureau of Labor Statistics (BLS) and the Census Bureau monitored the impact of this expansion on the CPS labor force estimates, and the BLS decided to begin using the expanded sample in its official CPS estimates in July 2001. Since the focus of this expansion was on estimates of children's health insurance coverage, the CPS ASEC sample was also increased by bringing back three rotation groups of households from the summer of the previous year and also asking the ASEC in the rotation groups in February and in April that were not part of the ASEC sample for the month of March. Interviewing in these rotation groups was limited to all minority households and White households with children under 18.

Support Files

Person Characteristics File

The Person Characteristics File (PCF) is a Census Bureau file that holds basic person-level descriptive data for all persons who have a Social Security Number (SSN). For each represented person, PCF holds summarized Social Security Administration (SSA) Numident data, consisting of a single record. In addition, the PCF holds a modeled race, ethnicity and gender for these same persons as well as Census 2000 reported race and ethnicity when available.

Master Address File Auxiliary Reference File

The Master Address File Auxiliary Reference File (MAF-ARF) is an annually-compiled Census Bureau file with only two fields: Master Address File Identifier (MAFID) and Protected Identification Key (PIK). The MAFID is the identity key for a Census Bureau's Master Address File (MAF) record and each represents a specific address. The PIK is a unique person-identifying code that replaces the Social Security Number to protect represented-persons' privacy. Each MAF-ARF record reflects that the person represented by the PIK is shown in the Census Administrative records collection associated (through residence or some other means) to the address represented by the MAFID.

Research Methods

Person Identification Validation System

The Person Identification Validation System (PVS), managed by the Administrative Records Applications staff, provides the Census Bureau with a fully-automated production capability for verifying or determining Social Security Numbers (SSNs) for person records within demographic surveys, censuses, or administrative records. This is accomplished by comparing person characteristics from the incoming file to the characteristics carried on the Census Numident file. In conformance with Census Bureau's privacy policy, the PVS does not process any record for which the respondent has refused to provide an SSN or has declined or "opted out" of having their data linked to administrative records from other agencies. For person records with reported SSNs, the system will attempt to verify the data associated with that SSN, and for records with no SSN or that fail this verification step, the system will attempt to determine the correct SSN through a probabilistic search. The SSN is considered validated if it successfully completed the verification step or is determined via search.

Adjusting the weights in the CPS to account for people missing SSN

Many sampled persons in the CPS are missing SSN because the respondent refused to provide it, or we were unable to identify it using the PVS. For this reason, we adjust the weights of persons for whom we have the SSN to account for those missing it, and remove the missings from the data file. The re-weighted data represent the whole survey frame.

The procedure is similar to the non-response weighting adjustment the Census Bureau and other survey researchers use. Appendix III Table 1 illustrates a non-response weighting adjustment by age and sex in a hypothetical survey of 1,000,000 persons (with known age and sex characteristics) and a sampling rate of 1 in 1,000. The number of sampled persons by age and sex is in Column 2. In the CPS, the population controls are by age, race, and sex for each state. Column 3 contains the number of sampled persons in each group responding to the survey. Response varies by group. For example, young males have an especially low response rate, as listed in Column 4.

The base weight of each sampled person, 1,000, is the inverse of the sampling rate. The procedure adjusts the base weight in Column 5 by the inverse of each group's response rate (Column 6) to produce the new weight in Column 7. Essentially, respondents with certain characteristics cover those with the same characteristics who didn't respond.

Note that multiplying the base weight of each group times the number of sampled persons and summing the products gives the number of people in the population, 1,000,000. Performing the same calculation with the new weight and the number of responses produces the same number, 1,000,000.

<u>1</u> Age/Sex Group	<u>2</u> Sampled Persons	<u>3</u> Responses	<u>4</u> Response Rate (Percent)	<u>5</u> Base Weight	<u>6</u> Adjustment Factor	<u>7</u> New Weight
YM	100	60	60.0	1,000	100/60	1,667
MM	250	188	75.0	1,000	100/75	1,333
OM	100	90	90.0	1,000	100/90	1,111
YF	100	80	80.0	1,000	100/80	1,250
MF	300	255	85.0	1,000	100/85	1,176
OF	150	142	95.0	1,000	100/95	1,053
Total	1,000	815	81.5			

Note: Groups are Young, Middle, Old; Male, Female. Sampling rate is 1 in 1,000 of 1,000,000-person population.

We use a similar procedure to adjust the weights of people in the CPS. Here, we are adjusting for missing SSN instead of missing response. For the CPS weighting adjustment, we calculate an adjustment factor in the same manner as the above example, by groups based on the variables most important to the research at hand, namely, age, relative poverty, health insurance status, and whether health insurance status was imputed.

- | | |
|---------------------------------------|--|
| Age | A 0-5
B 6-17
C 18-64
D 65 and up |
| Relative Poverty | A 0% to 99%
B 100% to 199%
C 200% and up |
| Health Insurance Status | 1 Medicaid
2 Other public only
3 Private only or other public/private
4 [not used]
5 Uninsured |
| Imputation of Health Insurance Status | I Imputed
N Nonimputed |

Note that in the above example, knowing the population's age and sex characteristics allows us to calculate adjustment factors for groups. In the CPS, we create groups based on persons' characteristics known because the Census Bureau collected (or imputed) the information in the survey.

Some of the groups created by crossing all of these variables were very small. We therefore collapsed them into other cells in order to bring the count (in each cell after the collapsing procedure) to at least 50. Specifically, in all age/poverty groups with imputed health insurance status, we combined *Other public only* with *Medicaid*. For the *Uninsured, 65 and up* group, we combined the poverty categories, and we combined the imputation categories.

After calculating the adjustment factors for the groups, we apply the appropriate factor to each sampled person's CPS weight, which the Census Bureau has already calculated for each person based on many characteristics. The adjustment factors, each the inverse of the SSN existence rate in the group, are in Appendix III Table 2.

Age Group	Relative Poverty	Health Insurance Status	Imputation of Health Insurance Status	Cell Count	Weighted number of persons (thousands)	Weighted number of persons with SSN (thousands)	Percent with SSN	Adjustment Factor
A	A	1	N	1,925	2,331	1,984	85.1	1.17
A	A	2	N	161	191	156	81.4	1.23
A	A	3	I	118	136	101	74.7	1.34
A	A	3	N	519	584	486	83.3	1.20
A	A	5	I	160	197	125	63.3	1.58
A	A	5	N	552	710	520	73.2	1.37
A	A	6	I	186	247	181	73.4	1.36
A	B	1	N	1,308	1,428	1,266	88.7	1.13
A	B	2	N	264	289	255	88.1	1.13
A	B	3	I	243	274	228	83.2	1.20
A	B	3	N	2,033	2,223	2,022	91.0	1.10
A	B	5	I	123	158	116	73.3	1.36
A	B	5	N	567	656	515	78.6	1.27
A	B	6	I	165	211	163	77.3	1.29
A	C	1	N	602	671	592	88.2	1.13
A	C	2	N	151	176	162	92.0	1.09
A	C	3	I	920	1,034	823	79.6	1.26
A	C	3	N	9,922	10,521	9,890	94.0	1.06
A	C	5	I	220	264	204	77.0	1.30
A	C	5	N	597	677	561	82.8	1.21
A	C	6	I	207	252	187	74.1	1.35
B	A	1	N	3,099	3,465	2,887	83.3	1.20
B	A	2	N	389	396	336	84.9	1.18
B	A	3	I	235	238	177	74.2	1.35
B	A	3	N	1,235	1,362	1,111	81.6	1.23
B	A	5	I	324	381	224	58.9	1.70
B	A	5	N	1,288	1,509	1,049	69.5	1.44
B	A	6	I	373	412	289	70.2	1.42
B	B	1	N	1,878	2,007	1,729	86.2	1.16
B	B	2	N	490	542	484	89.2	1.12
B	B	3	I	631	669	513	76.7	1.30
B	B	3	N	4,342	4,581	4,058	88.6	1.13
B	B	5	I	311	378	269	71.2	1.40
B	B	5	N	1,412	1,535	1,174	76.5	1.31
B	B	6	I	354	393	284	72.3	1.38
B	C	1	N	908	869	761	87.6	1.14
B	C	2	N	275	300	266	88.7	1.13
B	C	3	I	2,768	3,133	2,429	77.5	1.29
B	C	3	N	24,325	24,176	22,085	91.4	1.09

Appendix III Table 2. Weighting Adjustment Factors, March 2001 CPS Expanded Sample

Age Group	Relative Poverty	Health Insurance Status	Imputation of Health Insurance Status	Cell Count	Weighted number of persons (thousands)	Weighted number of persons with SSN (thousands)	Percent with SSN	Adjustment Factor
B	C	5	I	569	670	497	74.1	1.35
B	C	5	N	1,424	1,482	1,198	80.8	1.24
B	C	6	I	510	586	447	76.2	1.31
C	A	1	N	3,030	3,976	3,135	78.9	1.27
C	A	2	N	425	586	439	74.9	1.33
C	A	3	I	604	869	434	49.9	2.00
C	A	3	N	2,740	3,678	2,666	72.5	1.38
C	A	5	I	761	1,073	517	48.2	2.07
C	A	5	N	4,226	5,784	3,753	64.9	1.54
C	A	6	I	484	706	389	55.1	1.82
C	B	1	N	1,993	2,565	2,026	79.0	1.27
C	B	2	N	495	707	525	74.2	1.35
C	B	3	I	1,398	1,907	1,117	58.6	1.71
C	B	3	N	9,115	11,394	8,918	78.3	1.28
C	B	5	I	986	1,420	717	50.5	1.98
C	B	5	N	5,455	7,118	4,765	66.9	1.49
C	B	6	I	391	546	282	51.7	1.94
C	C	1	N	1,271	1,676	1,270	75.8	1.32
C	C	2	N	477	705	505	71.5	1.40
C	C	3	I	9,460	13,558	7,478	55.2	1.81
C	C	3	N	74,726	98,807	78,266	79.2	1.26
C	C	5	I	2,684	3,999	2,056	51.4	1.94
C	C	5	N	8,167	11,541	7,710	66.8	1.50
C	C	6	I	667	1,023	548	53.5	1.87
D	A	1	N	575	788	570	72.3	1.38
D	A	2	N	738	1,109	717	64.6	1.55
D	A	3	I	184	276	120	43.4	2.31
D	A	3	N	484	740	521	70.4	1.42
D	A	6	I	202	321	131	40.8	2.45
D	B	1	N	660	963	680	70.6	1.42
D	B	2	N	1,815	2,907	1,965	67.6	1.48
D	B	3	I	391	648	326	50.3	1.99
D	B	3	N	2,406	4,075	3,001	73.6	1.36
D	B	6	I	372	601	278	46.3	2.16
D	C	1	N	534	805	565	70.2	1.42
D	C	2	N	2,419	3,949	2,837	71.8	1.39
D	C	3	I	1,013	1,635	827	50.6	1.98
D	C	3	N	7,886	13,258	10,173	76.7	1.30
D	C	6	I	766	1,239	646	52.1	1.92
D	D	5	Z	186	251	133	52.9	1.89
Total				218,269	279,517	213,808	76.5	

Note: Health Insurance Status value 6 includes values 1 and 2. Imputation Status value Z includes values I and N.