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 Office of the Assistant Secretary for Planning and Evaluation, and the U.S. Census Bureau

---

Phase I Research Results:
Overview of National Medicare and Medicaid Files

Final
February 26, 2007

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 Center's for Medicare and Medicare Services (CMS), 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, CMS, or the US Census Bureau.

Contents

I. INTRODUCTION 1

II. OBJECTIVE AND SCOPE 1

III. QUESTIONS ANSWERED AND METHODOLOGY 2

IV. RESULTS 4

V. LIMITATIONS 15

VI. CONCLUSIONS 15

VII. APPENDICES 16

Appendix I-A: Description of Other Project Phases. 16

Appendix I-B: Description of Methods and Files Supporting Phase I Research 17

Appendix II: (Q1) MSIS CY 2000-2001 Charts and Tables by County and State .. 20

Appendix III: (Q2) MSIS Summary Tables, CY 2000-2001 20

Appendix IV: (Q3) MSIS Summary Tables, CY 2000-2001 20

Appendix V: (Q4) Joint (MSIS & EDB) Summary Tables 20

Appendix VI: (Q5) Joint (MSIS & EDB) Eligibility Status by State 20

Index of Tables

Table 1: Summary of MSIS versus MAFARF Address Coverage by Year 7
Table 2: MEDB (Medicare) Duplicate Records by Year..... 7
Table 3: MSIS (Medicaid) Duplicate Records by Year..... 8
Table 4: MSIS Versus EDB Aggregate Count of Dual-Eligible Individuals by State in 2001 8
Table 5: Race and Sex of Individuals Enrolled in Medicaid and Not Enrolled in for Medicare in 2001 12
Table 6: Race and Sex of Individuals Enrolled in Medicare and Enrolled in for Medicaid in 2001 12
Table 7: Race and Sex of Individuals Enrolled in both Medicare and Medicaid in 2001 13
Table 8: Medicaid Enrollees who are Not in Medicare by Ethnicity and Sex in 2001..... 13
Table 9: Medicare Enrollees who are Not in Medicaid by Ethnicity and Sex in 2001..... 14
Table 10: Dual Eligibles by Ethnicity and Sex in 2001..... 14

Index of Figures

Figure 1: U.S. Counties by their Rate of Validated SSNs in the MSIS in 2001..... 4
Figure 2: Frequency of States by their Rate of Validated SSNs in the MSIS in 2001 5
Figure 3: Frequency of Counties by their Rate of Validated SSNs in the MSIS in 2001... 5
Figure 4: U.S. States by their Rate of MSIS-EDB Agreement about Dual-Eligibility Status in 2001..... 11

I. Introduction

The paper describes the results of the first phase of a four-phase research project among the University of Minnesota's State Health Access Data Assistance Center (SHADAC), the Centers for Medicare and Medicaid Services (CMS), the U.S. Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation (ASPE), and the U.S. Census Bureau. The research project is designed to explain why discrepancies exist between Census Bureau survey estimates of enrollment in Medicaid and the number of enrollees reported in state and national administrative data. Project results will benefit the Census Bureau and other participating agencies because they can be used to improve evaluation of the Medicaid programs (e.g., estimating the effects of proposed policy changes) and to improve survey methods used to collect health insurance coverage information. The research project is divided into four phases; please see Appendix I-A for a description of each phase.

II. Objective and Scope

The objective of Phase I is to build a database of national health-insurance enrollment and to evaluate and summarize the data in it. We build the database by merging the CMS Medicaid Statistical Information System (MSIS) files with the CMS Medicare Enrollment Database (EDB) files. (See Appendix I-B for an explanation of the Social Security Number (SSN) validation process 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 characteristics of the U.S. population.

We assess our ability to accurately merge input files by analyzing the quality of the SSNs in the input files because SSNs are used to identify individuals and link records across different input files. A modified version of the Person Identification Validation System (PVS) is used to validate the SSNs on the input files against information from the Social Security Administration (SSA), and then a separate process is run to replace each SSN with a Protected Identification Key (PIK). Collectively these two functions allow record linkage while protecting personal privacy. (See Appendix I-B for a description of the PVS.)

We compare the database of health-insurance enrollment by producing descriptive statistics of person and address characteristics by program eligibility status for individuals who are: 1) enrolled only in Medicaid ("Medicaid-only"); 2) individuals who are enrolled only in Medicare ("Medicare-only"); and 3) individuals who are enrolled in both programs ("dual-eligible enrollees").

Phase I uses two Census Bureau files to assess the quality of the database of health-insurance enrollment: the Master Address File/Auxiliary Reference File (MAFARF), and the Person Characteristics File (PCF). (See Appendix I-B for a description of each supporting file.). The analysis covered calendar years (CY) 2000-2002; however,

because the results from each year were so similar, the main body of this report presents detailed results mostly for CY 2001. The details for all three years can be found in Appendices II-VI, which are provided separately on CD and are available upon request.

III. Questions Answered and Methodology

This section describes the research questions posed in this research and the methodology used to answer them.

Q1: What is the quality of the Social Security Number (SSN) fields on the EDB and the MSIS, and how well do the SSNs validate using the PVS? Is the quality of SSNs included in these files time variant across CY 2000, 2001 and 2002?

Use the following method for each input client[1] record from EDB and MSIS separately:

- Determine whether the original value in the SSN field falls within the SSA-defined acceptable range.
- After PVS processing, note whether a validated SSN is determined for the represented person.
- Produce descriptive statistics on the number of SSNs within the acceptable range and the number of SSNs validated by state and calendar year and by state, county, and calendar year.

Q2: Do the MSIS or EDB files contain geographic coding anomalies that could impact research results?

Separately for EDB for each year (CY 2000, 2001 and 2002) compare for each person with a validated SSN the CMS-coded state and county values to those existing in MAFARF for that same person for that year. Conclude that the EDB and MSIS state code matched if they agree with at least one of the MAFARF state codes for that person for that year. Similarly, conclude that EDB and MSIS county codes matched if both the CMS state and county codes agree with at least one MAFARF set of codes for that person for that year. After determining the state-level and county-level match status for each record, produce tallies for MSIS and EDB.

Q3: How many duplicate records exist in each file (MSIS and EDB) and how should these duplicate records be interpreted?

[1] We use the term "client" to refer to uniquely identified client accounts (as establish among records having the same State-MSIS Identification Number combination). Because in certain cases the same individual may have multiple client accounts, this use of the word "client" is different from the use of the word "person" or "individual." For any analysis done for clients, the MSIS data for a given calendar year is initially summarized to create a single record for each unique client account. For any analysis done for MSIS persons, summarization first creates a single record for each unique SSN. However, for clients whose SSNs cannot be validated, each is considered a single person or individual, since there is no reasonable basis for un-duplication.

Since an individual can be enrolled legitimately in one state for part of the month and then move to and enroll in another state during the same month, we assess the magnitude of duplication in enrollment records by creating two measures for this construct. One measure is a tally of the number of uniquely identified individuals with records in more than one state the same month and also within the same year. The other measure is a tally of the number of uniquely identified individuals with records showing more days of eligibility than days in a given month and year. We summarize these indicators of possible record duplication for each month and each year of the study period.

Q4: What is the universe of individuals who are defined as "dual eligible" because they were enrolled in both Medicare and Medicaid (i.e., there are records with PIKs in both the EDB and the MSIS)? What is the aggregate difference between the number of MSIS individuals who also appear in the EDB and the number of MSIS individuals with a dual-eligibility flag indicating they were enrolled in both Medicare and Medicaid? (Note: data from CY 2000 are excluded, as the MSIS dual-eligibility flag is unavailable for this year.)

To evaluate the consistency of information about dual-eligibility status we compare the Census Bureau indicator of dual eligibility to the CMS indicator of dual eligibility. We create an indicator of dual eligibility by matching to the EDB and using enrollment and termination dates both for Part A and Part B of the Medicare program. The CMS indicator of dual eligibility is the dual-eligibility flag on the MSIS record. We compare the aggregate counts using the CMS indicator to the counts using EDB information. We make the comparisons at the state level for each month of CY 2000 through CY 2002 and at the national and state levels for each calendar year.

Q5: How do the demographic and programmatic characteristics of individuals in the database of health-insurance enrollment compare to the demographics of the U.S. population and eligibility rules for the Medicare and Medicaid programs?

Merge the database of health-insurance enrollment with the PCF to supplement the database with additional person characteristics. Produce cross-tabulations by calendar year for individuals in the database according to the following:

- Program enrollment status (Medicaid-only, Medicare-only, or dual-eligible as of April 1st of the file year taken from EDB and MSIS)
- State of residence (based on CMS state code from EDB or alternatively from MSIS if person or valid state code not on EDB)
- Age (as of April 1 of the file year taken from PCF)
- Sex (from PCF)
- Race (from Census 2000 or modeled if not available)
- Ethnicity (from Census 2000 or modeled if not available)

Assess the reasonableness of the database information in terms of what we know about the types of individuals who should be in the database.

IV. Results

EDB validation rates were almost perfect, but MSIS validation rates were less so and varied significantly by state.

In 2001, EDB and MSIS Validation Rates were 99.8 and 89.4 percent, respectively. State-wise, MSIS validation rate ranged from a low of 67.3 for California to 99.0 percent for Vermont. Figure 1 maps county-level MSIS validation rates and Figures 2 and 3 show histograms of state and county validation results, respectively. Additionally, we provide an MSIS validation overview for 2001 that quantifies validation failure by type.

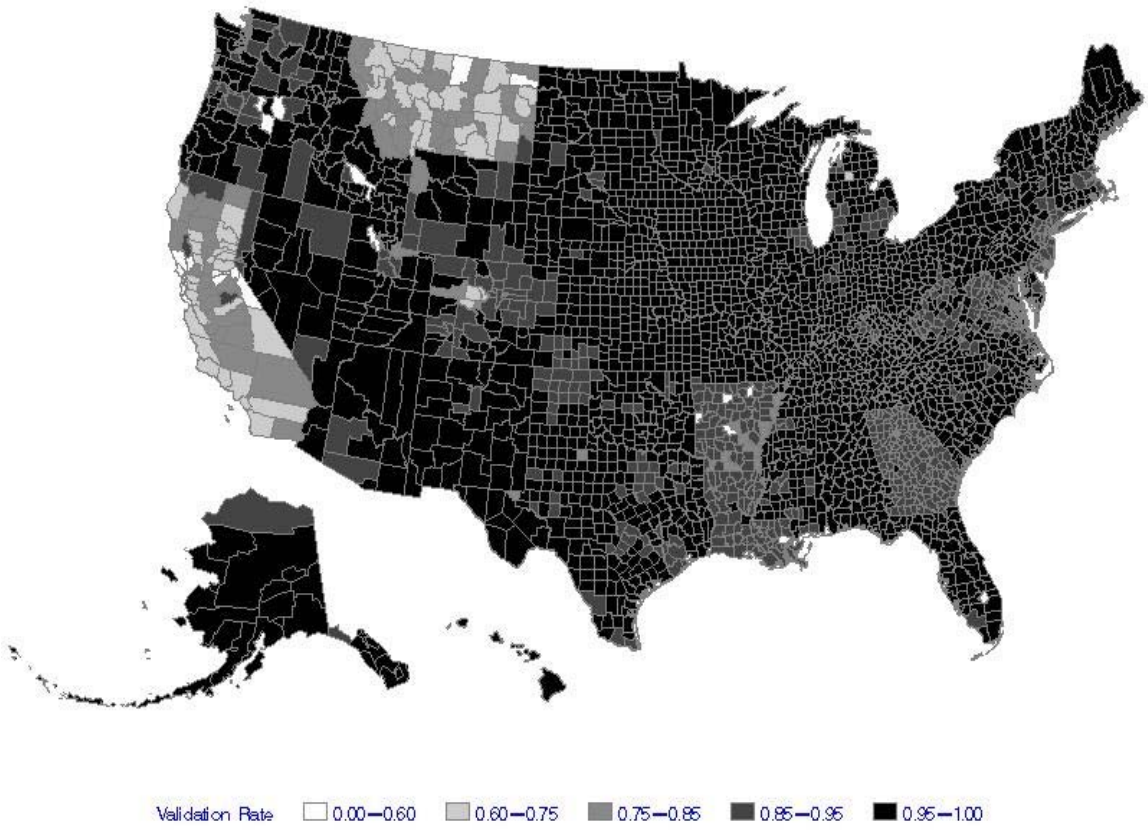

Figure 1: U.S. Counties by their Rate of Validated SSNs in the MSIS in 2001

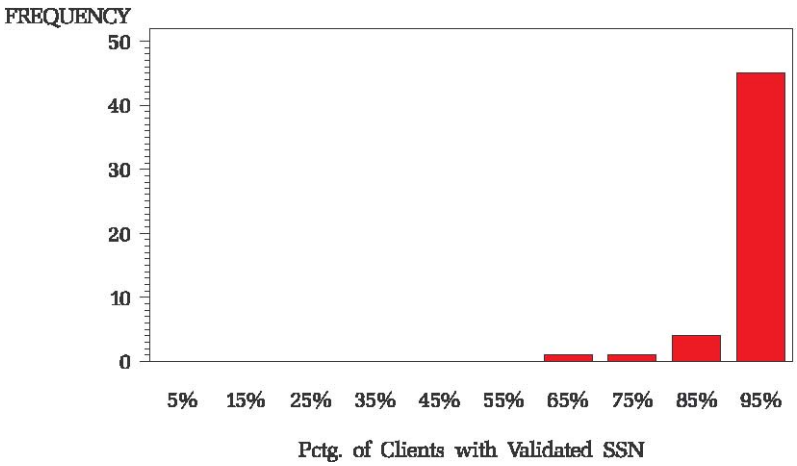

Figure 2: Frequency of States by their Rate of Validated SSNs in the MSIS in 2001

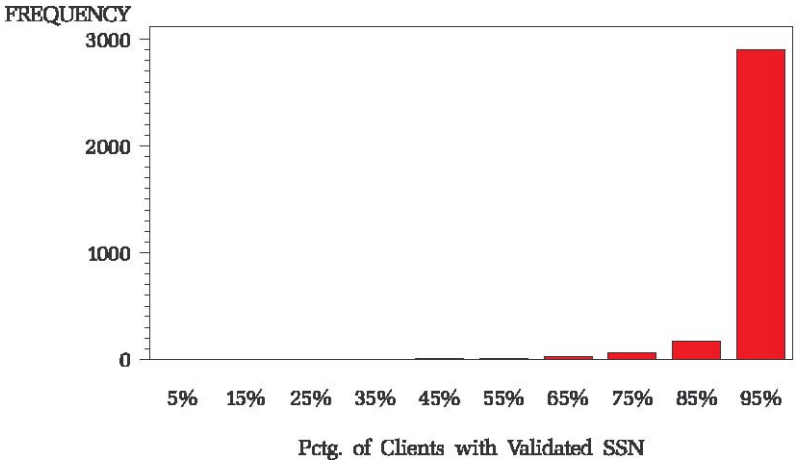

Figure 3: Frequency of Counties by their Rate of Validated SSNs in the MSIS in 2001

Overview of the 2001 MSIS SSN Validation Process:

Total 2001 MSIS Clients	49,272,652	
SSN Field is Blank (*pct of total clients*)	4,457,746	*(9.0%)*
SSN Contains Non-Numeric Values (*pct of total clients*)	18,098	*(0.0%)*
SSN in Taxpayer ID Number (TIN) Range (*pct of total clients*)	10,582	*(0.0%)*
SSN Greater than Numident[2] Maximum But Not in TIN Range (*pct.of total cl.*)	169,884	*(0.3%)*
SSN Otherwise Invalid (*pct of total clients*)	+10,492	*(0.0%)*
SSN is in Unacceptable Format (*pct of total clients*)	4,666,802	*(9.5%)*
SSN in Acceptable Format but not Found In Numident (*pct of total clients*)	+52,689	(*0.1%)*
SSNs Not in Numident (*pct of total clients*)	4,719,491	*(9.6%)*
Clients with SSN in Acceptable Range	44,553,161	
(*Pct of Total MSIS Clients*)	(*90.4%*)	
SSNs with unknown DOBs in Numident *(pct. of SSN in acceptable. range)*	5,943	*(0.0%s)*
SSNs Failing Validation *(pct. of SSN in acceptable range)*	+507,307	*(1.1%)*
SSNs in Acceptable Range not Validated *(pct. of SSN in acceptable. range)*	513,250	*(1.2%)*
Clients with Validated SSN	44,039,911	
(Pct. of SSN in acceptable range)	*(98.8%)*	
(Pct. of total MSIS clients)	*(89.4%)*	

The file resulting from this validation of MSIS becomes a requisite source file for the Phase II analysis. By having validated SSNs (subsequently converted to Protected Identity Keys—PIKs) a reliable match of MSIS data to CPS data can be made (by linking on the PIK). In each case where a CPS individual record can be linked to an MSIS record showing enrollment in Medicaid in the year of interest, we can know with near certainty that if the CPS data did not reflect this enrollment, then a 'false-negative' data coding error (i.e., survey measurement error) has occurred in the CPS.

MSIS geographic coding has anomalies that could affect subsequent research[3].

MSIS state coding agrees with internal Census Bureau administrative-record information (described in Appendix II) on residency 95.2 percent of the time. MSIS county coding agrees with the administrative records 83.5 percent of the time. This finding contrasts with findings from analyzing the EDB county coding because the EDB coding agree with the administrative records 95.1 percent of the time. However, this finding must be considered cautiously as only about 70 percent of MSIS clients were actually found in our administrative records[4].

[2] The Numident File is the Social Security Administration's (SSA) Numerical Identification File. It is a complete SSA transaction file for established Social Security Numbers (SSNs) and is indexed by SSN. If the value in the MSIS SSN field is greater than the maximum value for SSN on Numident, then it is necessarily invalid.

[3] See Appendix III, provided separately on CD and available upon request.

[4] Note that being "found in administrative records" is not the same as being validated. For validation, comparison is made specifically to the Census Numident, which is not considered by us (for policy reasons) to be an administrative records file, but rather as a reference file. Validation occurs at a significantly higher rate than being "found in administrative records" (at about 89% rather that 70%) because the Numident file is significantly more complete (as received by Census Bureau) and is unaffected by inability to geo-code addresses.

Table 1: Summary of MSIS versus MAFARF Address Coverage by Year

Year	Clients with Validated SSNs	Clients in MAFARF	Clients with MAFARF State Agreeing with MSIS State	Pctg. Clients with Matching MAFARF State	Clients with Valid MSIS County and in MAFARF	Clients with MAFARF County Agreeing with MSIS County	Pctg. Clients with Matching MAFARF County
2000	41,201,427	28,347,375	27,049,805	95.4%	28,199,081	23,680,866	84.0%
2001	44,039,911	31,193,690	29,687,560	95.2%	31,010,486	25,880,372	83.5%
2002	48,556,590	34,271,147	32,643,665	95.3%	33,039,460	27,709,283	83.9%

The EDB contains a negligible amount of duplicates, while MSIS has somewhat more (see Table 2 and Table 3 below).

Duplicate client records compose less than one-tenth percent of the EDB. About 2.2 percent of the MSIS uniquely identified individuals have records in multiple states simultaneously, suggesting possible duplication. However, since an individual can be enrolled legitimately in Medicaid in one state for part of the month and then move to and enroll in another state during the same month, we also tally the number of uniquely identified individuals with more eligible days than days in a given month. By this measure of possible duplicate enrollment, about 1.7 percent of the MSIS person records were duplicates.

Table 2: MEDB (Medicare) Duplicate Records by Year

Year	Total Records with Validated SSNs	Total Unique Validated SSNs	Validated SSNs with Only a Single Record	Validated SSNs with Duplicate Records	Validated SSNs with Two Records	Validated SSNs with Three or More Records
2000	40,934,532	40,912,357	40,890,210	22,147	22,119	28
2001	41,518,006	41,494,012	41,470,048	23,964	23,934	30
2002	42,857,486	42,832,722	42,807,987	24,735	24,706	29

Table 3: MSIS (Medicaid) Duplicate Records by Year

Year	Records with Validated SSNs*	Clients with Validated SSNs	Clients Sharing SSN in Excess of 1	Persons Not Eligible for Medicaid	Persons Eligible for Medicaid	Persons Eligible for More Than Number of Days Than in Period	Persons Eligible in Exactly One State	Persons Eligible in Exactly Two States	Persons Eligible in Exactly Three States	Persons Eligible in Four or More States
2000	411,192,207	41,201,427	1,289,926	408,804	39,502,697	704,790	38,572,022	897,953	30,830	1,892
2001	443,170,350	44,039,911	1,463,314	506,805	42,069,792	838,152	40,993,211	1036395	38,098	2,088
2002	484,683,087	48,556,590	2,087,188	573,942	45,895,460	932,745	44,725,710	1126381	40,969	2,400

* Each record represents a unique combination of MSIS Client ID and month.

Some discrepancy exists between the MSIS information about dual eligibility status and the information extracted from matching to the EDB (see Table 4 below).

Nationwide about 95 percent of the individuals identified by the MSIS indicator of dual eligibility were confirmed by EDB.

Table 4: MSIS Versus EDB
Aggregate Count[5] of Dual-Eligible Individuals by State in 2001

State	MSIS indicates individual is dual eligible	EDB indicates individual is dual eligible	Both MSIS and EDB indicate individual is dual eligible	Percent of MSIS dual eligibles confirmed in EDB	Percent of EDB dual eligibles Confirmed in MSIS
Alabama	144,277	145,269	140,479	97.4	96.7
Alaska	8,507	8,887	8,429	99.1	94.8
Arizona	60,196	58,128	57,648	95.8	99.2
Arkansas	96,886	77,930	77,141	79.6	99.0
California	842,208	823,040	813,855	96.6	98.9
Colorado	59,649	58,139	57,690	96.7	99.2
Connecticut	73,566	71,864	71,698	97.5	99.8
Delaware	13,083	13,529	12,774	97.6	94.4
District of Columbia	14,459	14,888	14,020	97.0	94.2

[5] The aggregate count is the mean value of the 12 monthly aggregate counts.

Table 4: MSIS Versus EDB
Aggregate Count of Dual-Eligible Individuals by State in 2001 (cont'd.)

State	MSIS indicates individual is dual eligible	EDB indicates individual is dual eligible	Both MSIS and EDB indicate individual is dual eligible	Percent of MSIS dual eligibles confirmed in EDB	Percent of EDB dual eligibles Confirmed in MSIS
Florida	356,356	365,332	348,915	97.9	95.5
Georgia	160,010	164,106	156,970	98.1	95.7
Hawaii	20,267	20,204	19,511	96.3	96.6
Idaho	11,296	17,424	11,131	98.5	63.9
Illinois	165,470	161,652	158,608	95.9	98.1
Indiana	104,655	103,874	102,625	98.1	98.8
Iowa	56,393	55,960	55,339	98.1	98.9
Kansas	41,739	42,519	40,736	97.6	95.8
Kentucky	151,762	123,284	119,537	78.8	97.0
Louisiana	123,638	125,673	121,329	98.1	96.5
Maine	61,930	65,725	61,061	98.6	92.9
Maryland	76,198	73,437	72,559	95.2	98.8
Massachusetts	191,091	186,055	185,153	96.9	99.5
Michigan	182,038	179,849	177,010	97.2	98.4
Minnesota	92,646	92,294	90,506	97.7	98.1
Mississippi	123,524	123,022	121,290	98.2	98.6
Missouri	130,926	131,757	128,889	98.4	97.8
Montana	13,255	14,733	12,903	97.3	87.6
More Than One	33,214	29,968	29,717	89.5	99.2
Nebraska	30,661	30,239	30,149	98.3	99.7
Nevada	23,012	21,714	21,149	91.9	97.4
New Hampshire	16,126	16,337	15,856	98.3	97.1
New Jersey	157,315	157,495	151,202	96.1	96.0
New Mexico	34,703	35,165	33,807	97.4	96.1
New York	517,009	502,336	494,377	95.6	98.4
North Carolina	241,979	235,584	233,974	96.7	99.3
North Dakota	12,133	12,028	11,953	98.5	99.4
Ohio	176,779	181,849	173,951	98.4	95.7

Table 4: MSIS Versus EDB
Aggregate Count of Dual-Eligible Individuals by State in 2001 (cont'd.)

State	MSIS indicates individual is dual eligible	EDB indicates individual is dual eligible	Both MSIS and EDB indicate individual is dual eligible	Percent of MSIS dual eligibles confirmed in EDB	Percent of EDB dual eligibles Confirmed in MSIS
Oklahoma	79,161	77,435	76,875	97.1	99.3
Oregon	66,854	59,544	59,033	88.3	99.1
Pennsylvania	266,502	260,386	257,471	96.6	98.9
Rhode Island	28,634	28,956	27,854	97.3	96.2
South Carolina	107,104	108,694	104,633	97.7	96.3
South Dakota	14,918	14,949	14,701	98.5	98.3
Tennessee	267,297	226,433	222,757	83.3	98.4
Texas	438,740	425,438	420,923	95.9	98.9
Utah	14,917	16,289	14,612	98.0	89.7
Vermont	24,775	25,360	24,374	98.4	96.1
Virginia	124,426	120,437	119,709	96.2	99.4
Washington	94,845	94,767	92,404	97.4	97.5
West Virginia	44,082	46,804	43,297	98.2	92.5
Wisconsin	104,382	103,158	102,255	98.0	99.1
Wyoming	7,029	6,981	6,925	98.5	99.2
Total U.S.	2,839,710	2,756,425	2,702,943	95.2	98.1

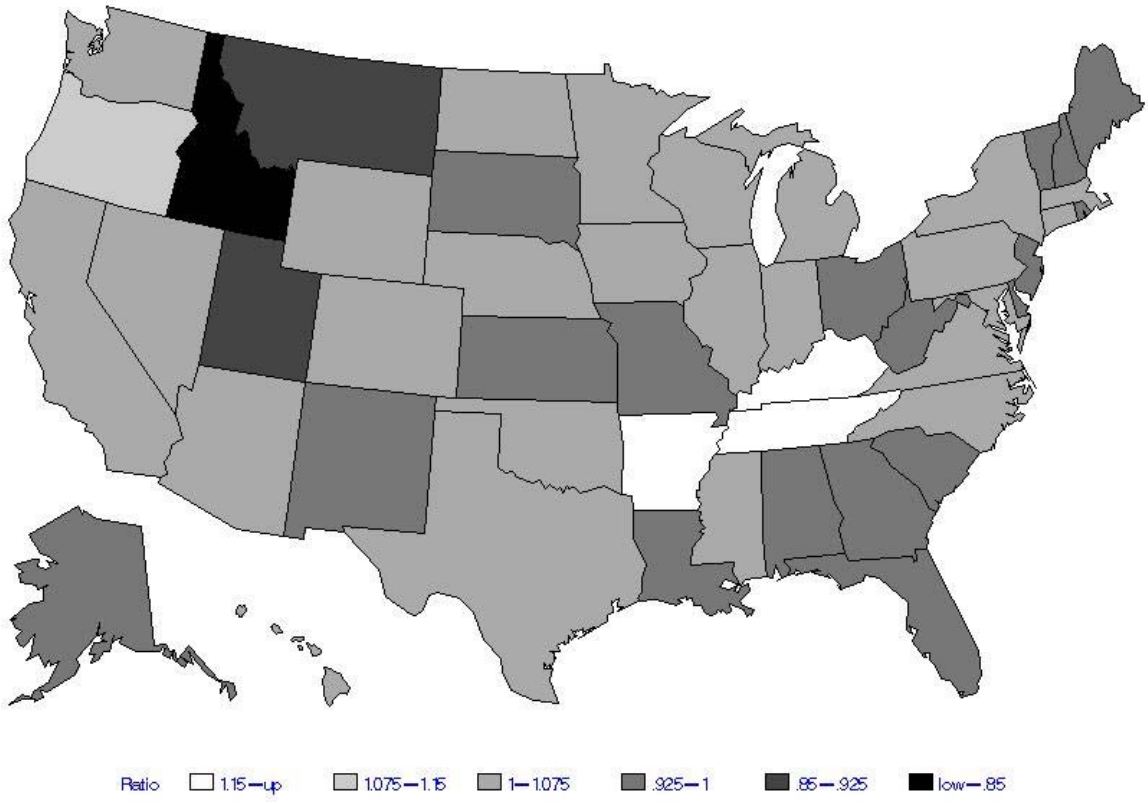

Figure 4: U.S. States by their Rate of MSIS-EDB Agreement about Dual-Eligibility Status in 2001

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[6].

Older white and black females comprise a disproportionate share of Medicare-only enrollees and dual-eligible enrollee persons, corresponding to women's disproportionate share of the broader older U.S. population. Younger white and black females comprise a disproportionate share of Medicaid-only enrollees. This corresponds to eligibility rules for Medicaid and Medicare. Tables 5, 6 and 7 show summary of the 2001 results by race and sex.

[6] See Appendix VI, provided separately on CD and available upon request.

Table 5: Race and Sex of Individuals Enrolled in Medicaid and Not Enrolled in Medicare in 2001

Race[7]	Sex	Count in MSIS	Percent of Total MSIS Count	Percent of Total U.S. Population[8]
Asian or Pacific Islander	Female	603,977	2.2	2.1
Asian or Pacific Islander	Male	594,159	2.2	2.0
Black	Female	4,647,562	16.9	6.7
Black	Male	3,592,524	13.1	6.0
American Indian or Alaskan Native	Female	300,056	1.1	0.5
American Indian or Alaskan Native	Male	243,593	0.9	0.5
White	Female	9,827,593	35.8	40.9
White	Male	7,674,665	27.9	39.9
Total		27,484,129	100.0	98.6

Table 6: Race and Sex of Individuals Enrolled in Medicare and Enrolled in Medicaid in 2001

Race	Sex	Count in EDB	Percent of Total EDB Count	Percent of Total U.S. Population
Asian or Pacific Islander	Female	318,695	0.9	2.1
Asian or Pacific Islander	Male	274,983	0.8	2.0
Black	Female	1,451,684	4.3	6.7
Black	Male	1,220,843	3.6	6.0
American Indian or Alaskan Native	Female	81,118	0.2	0.5
American Indian or Alaskan Native	Male	69,438	0.2	0.5
White	Female	16,932,826	49.8	40.9
White	Male	13,672,846	40.2	39.9
Total		34,022,433	100.0	98.6[9]

[7] Racial categories exclude individuals reported as having a multi-racial heritage.
[8] Census Bureau estimates of the U.S. resident population for July 1, 2001. For more information on the methodology used to produce these estimates see http://www.census.gov/popest/national/asrh/.
[9] See footnote 2 for explanation of why this column does not sum to 100.

Table 7: Race and Sex of Individuals Enrolled in both Medicare and Medicaid in 2001

Race	Sex	Count in both MSIS and EDB	Percent of Total Dual-Eligible Count	Percent of Total U.S. Population
Asian or Pacific Islander	Female	177,226	2.9	2.1
Asian or Pacific Islander.	Male	156,478	2.6	2.0
Black	Female	806,427	13.3	6.7
Black	Male	439,548	7.2	6.0
American Indian or Alaskan Native	Female	35,632	0.6	0.5
American Indian or Alaskan Native	Male	22,855	0.4	0.5
White	Female	2,822,574	46.4	40.9
White	Male	1,622,251	26.7	39.9
Total		6,082,991	100.0	98.6

One interesting finding is that Hispanics are disproportionately overrepresented in Medicaid and underrepresented in Medicare. Appendix VI shows that while about 12 percent of Medicaid enrollees are Hispanic females only 2.9 percent of Medicare enrollees are Hispanic females. This is consistent with the age distribution of Hispanic females in the general population—that is, the population of Hispanic females is younger than the non-Hispanic population so Hispanic females are less likely to be in Medicare.

An overview of the 2001 results by Ethnicity and sex are shown in Tables 8, 9 and 10.

Table 8: Medicaid Enrollees who are Not in Medicare by Ethnicity and Sex in 2001

Ethnicity	Sex	Frequency Count	Percent of Total Frequency	Percent of Total Medicaid Population
Hispanic	Female	3,266,926	11.9	6.3
Hispanic	Male	2,674,072	9.7	6.7
Non-Hispanic	Female	12,112,262	44.1	44.6
Non-Hispanic	Male	9,430,869	34.3	42.4
Total		27,484,129	100.0	100.0

Table 9: Medicare Enrollees who are Not in Medicaid by Ethnicity and Sex in 2001

Ethnicity	Sex	Frequency Count	Percent of Total Frequency	Percent of Total Medicare Population
Hispanic	Female	994,099	2.9	6.3
Hispanic	Male	938,219	2.8	6.7
Non-Hispanic	Female	17,790,224	52.3	44.6
Non-Hispanic	Male	14,299,891	42.0	42.4
Total		34,022,433	100.0	100.0

Table 10: Dual-Eligible Enrollees in Medicare and Medicaid by Ethnicity and Sex in 2001

Ethnicity	Sex	Frequency Count	Percent of Total Frequency	Percent of Total Dual-Eligible Population
Hispanic	Female	469,437	7.7	6.3
Hispanic	Male	308,077	5.1	6.7
Non-Hispanic	Female	3,372,422	55.4	44.6
Non-Hispanic	Male	1,933,055	31.8	42.4
Total		6,082,991	100.0	100.0

V. Limitations

We noted the following limitations in our research:

- The national MSIS file only contains SSN, date-of-birth and sex, a fact which constrains validation rates.
- Because only about 90% of MSIS clients have personal identity confirmed (i.e., their MSIS SSN value is validated), analyses that involve record linkage with MSIS may generate biased statistics:
    - Analysis of Questions 2, 4, and 5 may be affected marginally, as non-validated MSIS records may have different characteristics from validated ones.
    - The Question 3 analysis may significantly understate the rate of duplication within MSIS, since if one of the truly duplicated client accounts (i.e., from among multiple client accounts established for the same actual individual) cannot be assigned to a particular individual (due to not knowing the correct SSN), this fact of duplication cannot be established. This circumstance may have some impact on results developed in later phases of this project, particularly those measuring the size of the CPS estimate-MSIS count discrepancy.
- Since about 30% of MSIS individuals cannot be located on the MAFARF or Census Administrative data, the Question 2 analysis will be biased to the extent that non-MAFARF locatable persons are different from those that are locatable.

VI. Conclusions

Overall, the quality of the EDB exceeds the quality of the MSIS, in that the latter file seems to have more of a problem with duplicate and missing SSNs than EDB. No surprises emerged from the demographic analysis of the different program enrollees, including the dearth of Hispanics enrolled in Medicare. Our bottom line conclusion, however, is that the national 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. This work will be conducted as Phase II of the overall project.

VII. Appendices

Appendix I-A: Description of Other Project 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) files with the CMS Medicare (EDB) files. (See Appendix I-B for an explanation of the validation process and supporting files). We evaluate the quality of the database of health-insurance enrollment 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.

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

In Phase II we use data from the Medicaid Analytic Extract (MAX), MSIS, the Current Population Survey (CPS), the Person Characteristic File (PCF), and the Master Address File Auxiliary Reference File (MAFARF) (see Appendix I-B for an explanation of supporting files). Once matches/non-matches between the MSIS and CPS are determined, we supplement the matched records with information from the MAX and examine why there are discrepancies between MSIS records of enrollment and CPS reports of Medicaid coverage. The MAFARF and PCF are used to augment the analysis.

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

In Phase III we use data from the state MSIS files, CPS, 2001 Supplemental Survey (SS01), as well as the MAX and Census Bureau Master Address File (MAF). The Census Bureau works with CMS to negotiate the acquisition of state MSIS files that enhance the capability of finding matches because they include person names and addresses (in contrast with the national file, which does not include this identifying information). We compare the results of the Phase III analysis to the results from the Phase II analysis to determine if the expected higher match rate (given the additional information for finding matches in Phase III) affects the substantive conclusions of the analysis. We also analyze frame coverage to determine if frame differences (i.e., incomplete identification of individuals) affect the substantive conclusions of the analysis.

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

In Phase IV, we re-use 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 it.

Appendix I-B: Description of Methods and Files Supporting Phase I Research

Person ID Validation System (PVS)

The Person Identification Validation System (PVS) provides a fully automated production capability at the Census Bureau for Social Security Number (SSN) validation. Once an SSN is either verified or searched for and assigned, the record is considered validated. The PVS is managed by the Administrative Records Research Staff. The PVS enables SSN validation for regularly-repeating demographic surveys such as the CPS and SIPP, as well as for other demographic or administrative files. The PVS also expands data linkage capabilities for merging survey and administrative data sets.

The PVS uses probabilistic matching to verify SSNs contained within an incoming file against those contained within the Census Numident. The processing consists of a verification phase followed by a two-step search phase for assigning SSNs when necessary. For the verification phase, SSNs are matched using several types of demographic data, including names, dates of birth and gender. Specific weights are set to define acceptable matches. Any records not verified through this phase – or without an incoming SSN – are sent forward to the search phase of the system. (Note: In conformance with Census Bureau privacy policy, the PVS does not process any record for which the respondent has refused to provide his or her SSN. Also, due to technical constraints, the PVS does not process records where the respondent withholds both his or her first name and surname.)

The search phase of the PVS, also based on probabilistic matching, is comprised of a geokey (address-based) search, followed by a name search. The geokey search consists of logically grouping or "blocking" the data using the geokey, then progressively relaxing the geographic criteria while undertaking multiple passes through a matching routine to achieve agreement on demographic data as cited for the verification phase. Unmatched records remaining after the geokey search fall to the name search, where they undergo a similar demographic matching process but without the use of the geokey.

The final output file of the PVS (created after completion of the verification and the search phases) contains: all records with verified or searched and assigned SSNs; all records where the SSN could not be verified or searched and assigned or where multiple and, therefore, inconclusive SSNs were found; and, all original records withheld from the PVS process due to refusals or wholly blank names. A record is considered ***validated*** when it successfully completes either the verification or the assignment phase (geokey- or name-based search). Only validated records can be used in record linkage.

Master Address File (MAF)

The Master Address File is a Census Bureau File that contains an accurate, up to date inventory of all known living quarters in the United States, Puerto Rico and associated island areas. The MAF is used to support most of the census and surveys that the Census Bureau conducts including the decennial census, the American Community Survey and ongoing demographic surveys. The content of the MAF includes address information, Census geographic location codes, as well as source and history data.

Person Characteristics File (PCF)

The Person Characteristic File is a Census Bureau file that holds basic person-level descriptive data for all persons who hold a Social Security Number. The PCF is designed to provide a modeled race and gender for every person record present in the Census Numident regardless if race, ethnicity, or sex-specifying data is present.

Master Address File Auxiliary Reference File (MAFARF)

The Master Address File Auxiliary Reference File is a Census Bureau file that contains a Master Address File Identifier (MAFID) and a protected identification key (PIK). The MAFID is drawn from the MAF. The PIK is a resulting field from the verification of a record Social Security Number (SSN). At the completion of the SSN Verification process, the SSN is replaced with PIK as a form of anonymization of the record.

Medicare Enrollment Database File (EDB)

The Medicare Enrollment Database is the current CMS repository of enrollment and entitlement data for persons who are or have ever been enrolled in Medicare.

The EDB contains current and historical Medicare enrollment and entitlement information for all beneficiaries ever enrolled in the Hospital Insurance (HI) or Supplementary Medical Insurance (SMI) Medicare programs. The EDB replaced the Health Insurance Master (HIMA) File as the designated CMS repository of enrollment and entitlement data. As the primary source of information on demographic characteristics and geographic distribution of the entire Medicare population, the EDB supports various CMS and external research applications.

Program and policy analysts, as well as other researchers in health care related fields, use the person identifiers present on EDB records to construct samples and link data from a variety of sources at the beneficiary level. Analytic files derived from the EDB are used by CMS to support program-reporting requirements such as the production of statistical profiles of the Medicare population.

Beneficiaries are enrolled in Medicare based on criteria defined in Title XVIII of the Social Security Act of 1965 and subsequent amendments to the Act. EDB records are generated by the entitlement and enrollment of beneficiaries in the Medicare program.

The Social Security Administration (SSA) maintains demographic and entitlement information on all Medicare beneficiaries for whom entitlement is derived from Old Age and Survivors Insurance, disability insurance, End Stage Renal Disease (ESRD) programs, and the Railroad Retirement Board (RRB). The SSA Master Beneficiary Record (MBR) is the primary source of data for the EDB.

Beneficiary demographic characteristics, dates of enrollment and termination, cross-reference claim numbers, changes of address, and all other data contained in the SSA MBR system are transmitted from SSA to CMS to update the EDB. Additional information on RRB beneficiaries is contained on a beneficiary master record maintained by the RRB. Although entitlement data from the EDB typically update the Common Work File (CWF), a small number of data elements from the CWF update the EDB, such as some dates of death and Medicare secondary payer information.
The EDB was implemented in 1991. Currently, the EDB contains records for all Medicare beneficiaries ever entitled to HI or SMI. As such, creation of the EDB is a continuous process, with records of past enrollees kept in the database while new records are added as new cohorts of beneficiaries enroll.

Medicaid Statistical Information System (MSIS)

The Medicaid Statistical Information System 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.

Appendix II: (Q1) MSIS CY 2000-2001 Charts and Tables by County and State

[Note: Appendices II-VI are stored separately on CD and are available upon request.]

Appendix III: (Q2) MSIS Summary Tables, CY 2000-2001

[Note: Appendices II-VI are stored separately on CD and are available upon request.]

Appendix IV: (Q3) MSIS Summary Tables, CY 2000-2001

[Note: Appendices II-VI are stored separately on CD and are available upon request.]

Appendix V: (Q4) Joint (MSIS & EDB) Summary Tables

[Note: Appendices II-VI are stored separately on CD and are available upon request.]

Appendix VI: (Q5) Joint (MSIS & EDB) Eligibility Status by State

[Note: Appendices II-VI are stored separately on CD and are available upon request.]