Social Security Numbers in Medicaid Records: Reporting and Validity, 2009

Final Report

January 15, 2013

John L. Czajka
Shinu Verghese

MATHEMATICA
Policy Research

This page has been left blank for double-sided copying.

Social Security Numbers in Medicaid Records: Reporting and Validity, 2009

Final Report

January 15, 2013

John L. Czajka
Shinu Verghese

MATHEMATICA
Policy Research

This page has been left blank for double-sided copying.

This page has been left blank for double-sided copying.

CONTENTS

This page has been left blank for double-sided copying.

TABLES

ACRONYMS

BIC	Beneficiary identification code
CHIP	Children's Health Insurance Program
CMS	Centers for Medicare & Medicaid Services
FY	Fiscal year
HIC	Health insurance claim number
ID	Identification number
M-CHIP	CHIP implemented as a Medicaid expansion
MAX	Medicaid Analytic eXtract
MSIS	Medicaid Statistical Information System
MSIS ID	Medicaid ID
S-CHIP	CHIP implemented as a separate state program
SSA	Social Security Administration
SSN	Social Security number
TIN	Taxpayer identification number

This page has been left blank for double-sided copying.

EXECUTIVE SUMMARY

The Medicaid records that states submit to the Centers for Medicare & Medicaid Services (CMS) through the Medicaid Statistical Information System (MSIS) do not contain names and addresses, but they do contain Social Security numbers (SSNs). And, any attempt to link MSIS records—or the research data developed from MSIS records—to other databases must rely almost exclusively on these SSNs. The effectiveness of these linkages and the validity of any research based on these linked data are directly dependent on the quality of the SSNs recorded in the MSIS files.

Previous research with Medicaid enrollment data has documented how often SSNs are actually present in Medicaid records and how often the reported SSNs pass basic range checks. These studies have not been able to go the additional step of determining how often the SSNs that pass range checks also pass tests of validity at the individual level. In this study, we address that limitation, using data assembled by the Census Bureau to assess the validity of reported SSNs in recent MSIS data by age group and state of residence. This final report presents integrated findings on both the reporting and validity of MSIS SSNs at the state and national levels.

In our examination of SSN reporting at the national level, we defined the following three outcomes that can be discerned with MSIS data: (1) a reported SSN falls within the broad range of SSNs assigned by the Social Security Administration (SSA), (2) a reported SSN falls outside the valid range, or (3) no SSN is reported. Out of 55.0 million Medicaid enrollees during the fourth quarter of federal fiscal year (FY) 2009, the most recent quarter with complete data for all states, 50.8 million or 92.30 percent had SSNs that fell within the valid range. Another 13,000 or 0.02 percent had SSNs that fell outside the valid range, and 4.2 million or 7.68 percent had missing SSNs.

Reported SSNs in the valid range were more common in some age groups than others. SSNs in the valid range were reported by 99.38 percent of elderly enrollees, 87.96 percent of non-elderly adult enrollees, 96.24 percent of non-infant child enrollees, and 73.82 percent of infant enrollees. For enrollee records with missing age, virtually none—0.04 percent—had a reported SSN in the valid range.

The incidence of reported SSNs in the valid range at the national level reflects varying rates of SSN reporting at the state level. In a dozen states, the proportion of Medicaid enrollees reporting SSNs in the valid range was 99.0 percent or higher, but in California it was only 68.2 percent. For seven other states, the proportion was below 95 percent, although none was lower than 93 percent. The remaining states fell between 95 and 99 percent.

The information collected from applications for SSNs is stored in an SSA database called the Numident. Under an inter-agency agreement, the Census Bureau obtains regular updates to the Numident and has enhanced this database with information collected from its censuses, surveys, and other administrative records. We used the Census Bureau Numident for 2010 to validate the 50.8 million fourth quarter 2009 MSIS SSNs that were in the valid range.

The Census Bureau validation program applies the following four tests that involve comparing gender and the three components of the date of birth (year, month, and day) between the MSIS record and the Numident:

- Test 1: The year of birth differs by no more than one year, and either the month or day agrees exactly, and gender agrees.

- Test 2: The year of birth agrees and the month and day are inverted.

- Test 3: The month and day of birth agree and gender agrees.

- Test 4: The year, month, and day of birth agree.

An SSN is classified as valid if the MSIS record on which it appears passes any one of these tests.

The four tests can be performed only when the MSIS SSN is found in the Numident. Therefore, there are three possible outcomes of the Census Bureau validation procedure: (1) an SSN passes validation, (2) an SSN is found in the Numident but fails validation, and (3) an SSN is not found in the Numident. We applied the validation tests to SSNs that fell within the valid range, as defined above. Of the 50.8 million records submitted for validation, 99.37 percent passed. Another 0.53 percent of the records had SSNs and were found in the Numident but failed all four of the tests. The SSNs on the remaining 0.10 percent of the records were not in the Numident.

While agreement on gender and date of birth did not have to be perfect for an SSN to be considered valid, it nearly always was. Across all age groups, 97.55 percent of the valid SSNs were on MSIS records with the same gender and date of birth as the corresponding Numident record. Another 1.29 percent agreed on all but the month or day of birth, and an additional 0.27 percent disagreed only on the year of birth, which differed by one. Together, these patterns accounted for 99.11 percent of the valid SSNs. When MSIS SSNs failed validation, the Numident and MSIS record disagreed on at least three elements among gender, month, day, and year of birth more than three-quarters of the time.

Combining the results on reporting and validation, we find the following. Valid SSNs were present on 91.7 percent of the MSIS records submitted by the states for the fourth quarter of 2009. SSNs were missing on 7.7 percent of the MSIS records and incorrect on another 0.6 percent of the records. When an SSN was reported, it was almost always correct: 99.3 percent of the reported SSNs were judged to be valid. With such a high validation rate overall, there was little variation among the states. Reported SSNs were valid at least 99.0 percent of the time in all but seven states and at least 98.5 percent in all but one state.

The exceedingly high validation rate for MSIS SSNs is significant to CMS because the validation results for individual Medicaid enrollees cannot be removed from the Census Bureau. With such a high validation rate, not having access to the validation results at the micro level may not matter. It is difficult to imagine an application where serious error could result from simply assuming that all reported SSNs are correct. For applications involving record linkage, use of an appropriate range check prior to linkage would eliminate most of the illegitimate SSNs (those in ranges that SSA has never assigned), and matching on gender and date of birth in addition to SSN would prevent nearly all matches to MSIS records with incorrect SSNs.

I. INTRODUCTION

Used as a personal identifier in a variety of administrative record systems and even collected in some surveys, the Social Security number (SSN) provides an instrument for linking the records of individuals across different databases. Without it, analyses that depend on linked data simply would not be feasible in many cases.

The Medicaid records that states submit to the federal government through the Medicaid Statistical Information System (MSIS) do not contain names and addresses, but they do contain SSNs.[1] Any attempt to link MSIS records—or the research data that are developed from MSIS records—to other databases must rely almost exclusively on these SSNs. The effectiveness of these linkages and the validity of any research based on these linked data are directly dependent on the quality of the SSNs that are recorded in the MSIS files.

Previous research with Medicaid enrollment data, discussed below, has documented how often SSNs are actually present in Medicaid records and how often the reported SSNs pass basic range checks. These studies have not been able to go the additional step of determining how often the SSNs that pass range checks also pass tests of validity at the individual level. This study addresses this limitation, using data assembled by the Census Bureau to assess the validity of reported SSNs in recent MSIS data by age group and state of residence. This final report presents integrated findings on both reporting and validity of MSIS SSNs at the state and national levels.

This first chapter provides an overview of issues followed by a brief review of prior research. Chapter II presents our empirical findings, and Chapter III discusses their implications and presents our final conclusions.

---

[1] Names will be captured in transformed MSIS (T-MSIS), which is scheduled to debut in January 2014.

A. Overview of Issues

The SSN is assigned by the Social Security Administration (SSA), and it is unique to the person to whom it is assigned. No two individuals are assigned the same SSN, so, in theory, a given SSN should correspond to only one individual.[2] However, the SSN was never intended to serve as a personal identifier; it was devised solely to track entitlements to retirement and other benefits provided under the social security system. The use of SSNs as a personal identifier by both businesses and government agencies has imposed demands that the SSN was not designed to address. For example, the SSN lacks check digits—common to credit card numbers—that would enable erroneous numbers to be identified easily. Also, the Social Security card itself is not suitable for identification purposes, so it is rarely required to be shown, making it easier for someone to supply an erroneous number—or to use another person's SSN.

The Centers for Medicare & Medicaid Services (CMS) encourages but does not require states to include SSNs in their MSIS submissions for all enrollees. Nor does CMS mandate that the states validate the SSNs that they do report. Consequently, SSNs may be missing or incorrect on individual MSIS records. Some errors may be corrected over time, but unless the corrections are applied retroactively, errors will persist in the MSIS records prior to the time that the corrections are implemented. One example is notable. If an SSN is unavailable—for instance, not yet assigned—at the time of enrollment, a state may assign a pseudo (or temporary) value to the SSN and later replace it with the true (or permanent) value. This is common with infants, who do not acquire SSNs until at least a few weeks after birth at the earliest. For this

---

[2] There may be circumstances—for example, proven identity theft—where an individual may be issued a new SSN, but such occurrences are so rare that their impact on an analysis such as the one presented here would be entirely negligible.

reason, it is important to separate infants from older children when conducting a validation study of Medicaid SSNs.

For persons dually enrolled in Medicaid and Medicare, the MSIS record provides a second identifier—the Medicare health insurance claim number (HIC). The HIC is an 11-digit number consisting of two parts. The first nine digits are the primary claimant's SSN, and the final two digits are the beneficiary identification code (BIC), which identifies the beneficiary's relationship to the primary claimant—most commonly self or "aged wife."

In addition to its limited availability (dual eligibles represent a small proportion of Medicaid enrollees in most if not all states and nationally—under 15 percent), the HIC has other limitations as a personal identifier (Czajka et al. 2010). As with the SSN, states are not required to verify the accuracy of the HIC reported in MSIS. Further, a HIC can change over time when the beneficiary's relationship changes—for example, a wife is divorced from or widowed by the primary claimant. Most states are not equipped to update the HIC in a timely manner, so the MSIS HIC may become out of sync with the one in the Medicare enrollment database. Lastly, the HIC is not nearly as widely used as the SSN, so its value as a linkage variable for matching Medicaid or Medicare records to other databases is very limited. For these reasons, assessments of the quality of the personal identifiers in MSIS tend to focus on the SSN.

B. Prior Research

We can express the quality of SSNs in Medicaid enrollment data in terms of two dimensions: (1) whether SSNs are reported at all and (2) whether those that are reported are recorded correctly. Previous research has examined the frequency and characteristics of MSIS records that lack SSNs, but except for broad range checks, the accuracy of the SSNs recorded in MSIS files has not been assessed.

An analysis of SSN quality in the Medicaid Analytic eXtract (MAX), a research file derived from MSIS data, found the following. Out of 60.6 million enrollees with eligibility records in the 2005 calendar year MAX file, 9.3 percent had no SSN, and 0.7 percent had an SSN that failed a basic test of validity—the high group test (Czajka et al. 2010). For SSNs issued prior to June 25, 2011, when SSA changed aspects of the format, the SSN had the form AAA-GG-SSSS, where AAA was a geographic area code, GG was a group code, and SSSS was a serial number. Within an area, both the group code and the serial number were assigned sequentially, with the serial number varying most rapidly. That is, all 9,999 serial numbers (0000 is not used) were assigned before the group code was incremented. Each month SSA published on its website the maximum group code used for each of the 772 areas.[3] No SSN with a higher group code had been issued, so the high group test compared the area and group codes on each SSN with the published numbers to determine if an SSN with these codes could have been issued. The high group test also checked for occurrences of 000 in the area code and 00 in the group code, as these values were not used. SSNs that failed the high group test were deemed invalid.[4]

Nationally, the frequency of missing SSNs was strongly influenced by California, which lacked SSNs for 35.7 percent of its eligibility records. California operates a large family planning program under Medicaid, and participants are not required to provide SSNs. While no state matched California in the relative frequency of missing SSNs, 11 other states were missing SSNs for at least 5 percent of their records, and only 10 states were missing SSNs for less than 1 percent of their records.

---

[3] For SSNs issued after June 24, 2011, the area code no longer has a geographic interpretation, and there is no longer a high group code associated with each area (see http://www.ssa.gov/employer/randomization.html). With this change, SSA has discontinued updates to the high group list. Previous lists available on the SSA website can be used for SSNs issued prior to June 25, 2011.

[4] The high group test provided no information about the validity of SSNs with group codes that had been issued.

There was state variability as well in the proportion of records that failed the high group test, although it was more constrained. Arizona (8.9 percent) and Montana (8.0 percent) had the highest failure rates followed by Arkansas (4.4 percent) and Hawaii (2.1 percent). Only three other states had high group test failure rates above 1 percent. Prior to 2005, Louisiana and Nevada had assigned an out-of-range pseudo SSN to records with missing SSNs, which produced high failure rates. Both states were asked to suspend this practice, which reduced their high group failure rates below 1 percent (but raised their missing SSN rates).

Duplicate SSNs within the same state were also observed among 0.6 percent of the records.[5] In four states—Illinois, New York, North Dakota, and South Dakota—the rate of duplicate SSNs exceeded 3 percent. In every other state but two the rate was below 1 percent and was often zero.

Czajka and Libersky (2011) examined the characteristics of the 9.91 percent of 2007 MAX records that were missing SSNs.[6] Table I.1 provides a breakdown of the nearly six million records with missing SSNs by benefit eligibility and age. Separate estimates for California and the balance of states are reported for those subpopulations for which California accounted for a disproportionate share of the enrollees nationally. The subpopulations are arrayed from largest to smallest in terms of the number of enrollees with missing SSNs. For each subpopulation, the table reports the number of enrollees with a missing SSN, the subpopulation's share of missing SSNs nationally, the total enrollees (with and without SSNs), and the percent of enrollees with a missing SSN.

---

[5] Each occurrence of a duplicate SSN was associated with a different MSIS ID, as MSIS IDs are supposed to be unique within each state in MAX, according to MSIS reporting instructions. Each occurrence of an SSN that appeared more than once was counted as a duplicate, although one occurrence was probably correct.

[6] This estimate excludes (from both the numerator and the denominator) more than 700,000 records that lacked dates of birth along with SSNs. Most of these were claims records that contained no eligibility information.

Table I.1. Distribution of Missing SSNs by Subpopulation and State Group, 2007

Subpopulation	State(s)	Number with a Missing SSN	Share of Total	Cumu- lative Share	Total Enrollees	Percent with a Missing SSN
Total		5,998,695	100.00		60,504,056	9.91
Family planning, ages 19 to 44	California	2,118,015	35.31	35.31	2,118,015	100.00
Infants with full benefits	All states	825,126	13.76	49.07	2,399,799	34.38
Children 1 to 5 with full benefits	All states	670,973	11.19	60.26	9,877,177	6.79
Aliens 19 to 44	California	630,736	10.51	70.77	708,008	89.09
Children 6 to 18 with restricted benefits	California	416,810	6.95	77.72	425,490	97.96
Children 6 to 18 with full benefits	All states	389,253	6.49	84.21	17,276,308	2.25
Aliens 19 to 44	All but CA	273,554	4.56	88.77	333,306	82.07
Adults 19 to 44 with full benefits	All states	246,834	4.11	92.88	12,987,019	1.90
Family planning, ages 45 to 64	California	134,316	2.24	95.12	134,316	100.00
Children 6 to 18 with restricted benefits	All but CA	72,939	1.22	96.34	131,053	55.66
Children with S-CHIP only	All states	62,528	1.04	97.38	934,016	6.69
Aliens 45 to 64	California	39,448	0.66	98.04	53,984	73.07
Children 1 to 5 with restricted benefits	All states	30,139	0.50	98.54	32,606	92.43
Adults 45 to 64 with full benefits	All states	25,163	0.42	98.96	5,637,701	0.45
Aliens 65 and older	California	18,554	0.31	99.27	23,358	79.43
Aliens 65 and older	All but CA	15,059	0.25	99.52	28,389	53.05
Adults 65 and older with full benefits	All states	8,999	0.15	99.67	4,717,471	0.19
Aliens 45 to 64	All but CA	7,496	0.12	99.79	12,419	60.36
Adults 19 to 44 with other restricted benefits	All states	5,671	0.09	99.88	171,041	3.32
Family planning, ages 19 to 44	All but CA	4,947	0.08	99.96	766,742	0.65
Infants with restricted benefits	All states	1,543	0.03	99.99	2,431	63.47
Adults 65 and older with other restricted benefits	All states	327	0.01	100.00	1,215,836	0.03
Adults 45 to 64 with other restricted benefits	All states	173	0.00	100.00	497,391	0.03
Family planning, ages 45 to 64	All but CA	92	0.00	100.00	20,180	0.46

Source: Czajka and Libersky (2011), from MAX PS files for calendar year 2007.

More than one-third (35.31 percent) of the total number of missing SSNs for all states was due to a single subpopulation in California—family planning enrollees 19 to 44, none of whom had an SSN. Infants with full benefits accounted for the next largest number of missing SSNs, followed by children ages 1 to 5 with full benefits and adults ages 19 to 44 with alien benefits in California. Together these four subpopulations accounted for 70.77 percent of the missing SSNs. Adding children ages 6 to 18 with either full or restricted benefits and aliens ages 19 to 44 in the rest of the states raised the cumulative share to 88.77 percent. The next five subpopulations

increased the cumulative share to 98.04 percent. The final two percent was due to 12 subpopulations drawn from the balance of enrollees.

Most of the subpopulations that contributed the largest numbers of enrollees with missing SSNs had very high rates of missing SSNs. For example, no SSNs were present for family planning participants in California, although this was a peculiarity of that state, as nearly all family planning participants in the rest of the country had SSNs (only 0.65 percent of the numbers were missing for family planning, ages 19 to 44, and 0.46 percent for family planning, ages 45 to 64). Also in California, SSNs were missing for 97.96 percent of the children ages 6 to 18 with restricted benefits, but this was part of a more general pattern, as SSNs were missing for 55.66 percent of the children with restricted benefits in other states as well. Similarly, SSNs were missing for 89.09 percent of the aliens ages 19 to 44 in California and 82.07 percent of this subpopulation elsewhere. However, children ages 1 to 5 with full benefits had a comparatively low rate of missing SSNs (6.79 percent) and yet were the third largest source of missing SSNs. Infants with full benefits had a missing rate of 34.38 percent and were the second largest contributor to the records with missing SSNs.

Even low rates of missing SSNs could produce large numbers of records with missing SSNs if the subpopulations were large enough. The two largest subpopulations—children ages 6 to 18 with full benefits and adults ages 19 to 44 with full benefits—had missing rates around 2 percent, but when combined with their overall size (17.2 million and 13.0 million, respectively) these rates produced some of the largest numbers of missing SSNs. By contrast, adults ages 45 to 64 and age 65 and older with full benefits were large subpopulations as well, numbering 5.6 million and 4.7 million, but their rates of missing SSNs were so low (0.45 percent and 0.19 percent) that their contribution to the total number of records with missing SSNs was negligible.

This page has been left blank for double-sided copying.

II. REPORTING AND VALIDITY OF SOCIAL SECURITY NUMBERS

States submit their data to MSIS on a quarterly basis. At this writing, the most recent quarter with complete MSIS data for all 50 states and the District of Columbia is the final quarter of the 2009 federal fiscal year (FY), or July through September 2009. To present the most current estimates of SSN validity, therefore, we base our analysis on data from the fourth quarter, FY 2009. For this period we examine both the reporting of SSNs and the validity of the reported numbers. As we will show, most of the variation in the quality of the SSNs in MSIS data is due to variation in the collection of SSNs; when numbers are reported, the vast majority appears to be valid. Section A presents findings on the reporting of SSNs, and Section B presents findings on the validity of reported SSNs.

We are also interested in whether the quality of SSNs has improved in recent years. To answer this question, we compared the estimates from the fourth quarter, FY 2009, with estimates from about three years earlier—or, more specifically, the first quarter, FY 2007, covering the months of October through December 2006. Findings from this comparison are presented in Section C.

A. Reporting of SSNs: Fourth Quarter, FY 2009

In our examination of SSN reporting at the national level, we define three outcomes that can be discerned with MSIS data: (1) a reported SSN falls within the broad range of SSNs assigned by the SSA, (2) a reported SSN falls outside the valid range, or (3) no SSN is reported.

For an SSN to be counted in the valid range, the first 3 digits (out of 9) must be no greater than 772, and the overall number must be no less than 001010001. In addition, each of the three segments—that is, the first three digits, the next two, and the final four—must be greater than zero regardless of the values in the other segments. We note that numbers beginning with 9 (that is, with the first 3 digits equal to 900 or more) are assigned as Taxpayer Identification Numbers

9

(TINs) to persons who file or are claimed as dependents on U.S. tax returns but do not qualify to receive SSNs.[7] Although TINS are legitimate numbers for tax purposes, they do not substitute for SSNs in MSIS. Further, TINS cannot be validated by SSA or the Census Bureau because neither agency possesses the information needed to do so. For these reasons we do not regard a number that could be a TIN as a valid SSN.

Reported SSNs with zeroes for any of the three segments (which includes all numbers less than 001010001) or with a number greater than 772 in the first segment are considered outside the valid range. Missing SSNs may be represented in different ways in MSIS but apparently not as blanks because we observed no such values—or any other non-numeric characters. We counted an SSN as missing if the field was filled with the same digit (other than zero, which falls under our definition of invalid) or one of the strings 0 through 8, 1 through 9, or 9 through 1 (which is not assigned as a TIN). Empirical findings are presented first for the nation and then by state.

1. National Results

Distributions of the three outcomes, by enrollee age, are reported for the nation in Table II.1.[8] Out of 55.0 million enrollees during the fourth quarter of FY2009, 50.8 million or 92.30 percent had SSNs that fell within the valid range. Another 13,000 or 0.02 percent had SSNs that fell outside the valid range, and 4.2 million or 7.68 percent had missing SSNs.

---

[7] Generally, TINs are reserved for foreign nationals.

[8] Medicaid enrollees include participants in the Children's Health Insurance Program (CHIP) when the latter is operated through a Medicaid expansion (M-CHIP). States were allowed to implement CHIP through a separate state program (S-CHIP) that was not required to offer the Medicaid benefit package, and a majority of states did so. The reporting of S-CHIP enrollment in MSIS is optional, however. About half of the states report S-CHIP enrollment in MSIS, but these states have only about a quarter of the total S-CHIP enrollment. For this reason, the estimates presented in this report are limited to persons with one or more months of Medicaid enrollment (including M-CHIP) during the reference period. MSIS records for persons who were enrolled in S-CHIP in all months of enrollment during the quarter are excluded from all analyses.

Table II.1. Total Medicaid Enrollees by Age Group, Benefit Type, and Presence of an SSN, Fourth Quarter, FY 2009

Description of SSN	Infants (< 1)	All Other Children (1 to 18)	Non-elderly Adults (19 to 64)	Elderly Adults (65+)	Records with Missing Age	Total
All Enrollees						
Total	2,341,788	26,055,073	21,165,392	5,399,709	65,587	55,027,549
SSN is in the valid range	1,728,738	25,076,500	18,617,842	5,366,111	27	50,789,218
SSN is outside the valid range[a]	1,157	9,845	1,947	133	0	13,082
SSN is missing[b]	611,893	968,728	2,545,603	33,465	65,560	4,225,249
Enrollees with Full Benefits in Any Month						
Total	2,340,721	25,551,352	17,007,964	4,157,241	65,573	49,122,851
SSN is in the valid range	1,728,275	25,007,598	16,891,675	4,146,814	27	47,774,389
SSN is outside the valid range[a]	1,157	9,758	1,337	86	0	12,338
SSN is missing[b]	611,289	533,996	114,952	10,341	65,546	1,336,124
Enrollees with Only Restricted Benefits						
Total	1,067	503,721	4,157,428	1,242,468	14	5,904,698
SSN is in the valid range	463	68,902	1,726,167	1,219,297	0	3,014,829
SSN is outside the valid range[a]	0	87	610	47	0	744
SSN is missing[b]	604	434,732	2,430,651	23,124	14	2,889,125
Percentage Distribution: All Enrollees						
SSN is in the valid range	73.82	96.24	87.96	99.38	0.04	92.30
SSN is outside the valid range[a]	0.05	0.04	0.01	0.00	0.00	0.02
SSN is missing[b]	26.13	3.72	12.03	0.62	99.96	7.68
Percentage Distribution: Enrollees with Full Benefits						
SSN is in the valid range	73.84	97.87	99.32	99.75	0.04	97.25
SSN is outside the valid range[a]	0.05	0.04	0.01	0.00	0.00	0.03
SSN is missing[b]	26.12	2.09	0.68	0.25	99.96	2.72
Percentage Distribution: Enrollees with Restricted Benefits						
SSN is in the valid range	43.39	13.68	41.52	98.14	0.00	51.06
SSN is outside the valid range[a]	0.00	0.02	0.01	0.00	0.00	0.01
SSN is missing[b]	56.61	86.30	58.47	1.86	100.00	48.93

Source: State MSIS files, fourth quarter, FY 2009.

[a] First three digits are greater than 772, or SSN is less than 001010001. Also includes SSNs with zeroes in the first three, next two, or final four digits.
[b] Includes SSN fields filled with the same digit (other than zero) or the strings 0 through 8, 1 through 9, or 9 through 1.

Reported SSNs in the valid range were more common in some age groups of enrollees than others. SSNs in the valid range were reported by 99.38 percent of elderly enrollees, 87.96 percent of non-elderly adult enrollees, 96.24 percent of non-infant child enrollees, and 73.82 percent of infant enrollees. For enrollee records with missing age, virtually none—0.04 percent—had a reported SSN in the valid range.

Overall and in every age group, reported SSNs in the valid range were more common among enrollees with full benefits than with restricted benefits. For enrollees with full benefits in any of the three months, 97.25 percent had SSNs in the valid range compared to 51.06 percent for enrollees with restricted benefits in all months of enrollment. Except for infants, enrollees with full benefits reported SSNs in the valid range between 97 and 100 percent of the time. For infants, this figure was 73.84 percent. Among enrollees with restricted benefits, however, only the elderly had a high rate of reporting of SSNs in the valid range; for this group, 98.14 percent reported SSNs in the valid range. For infants, 43.39 percent of those with restricted benefits (only 1,067 enrollees) reported SSNs in the valid range. Non-elderly adults were similar, with 41.52 percent reporting SSNs in the valid range, but this group accounted for nearly 4.2 million enrollees. Non-infant children had the lowest rate of reporting SSNs in the valid range at only 13.68 percent, but this group accounted for just over half a million enrollees. Of the more than 4.2 million enrollees with either a missing SSN or one that fell outside the valid range, 2.4 million or more than 57 percent were non-elderly adults with restricted benefits.

2. State Results

The incidence of reported SSNs in the valid range at the national level reflects varying rates of SSN reporting at the state level. In a dozen states, the proportion of Medicaid enrollees reporting SSNs in the valid range was 99.0 percent or higher, but in California it was only 68.2

percent (Table II.2). For seven other states, the proportion was below 95 percent, although none was lower than 93 percent. The remaining states fell between 95 and 99 percent.

For elderly enrollees the reporting rate for SSNs in the valid range was 100 percent in 24 states and within half a percent of that figure in all but five states. Massachusetts was lowest at 96.7 percent, and the remaining four states fell between 98.2 and 98.7 percent. We suspect that the nearly universal reporting of SSNs among the elderly reflects a combination of two factors: (1) the widespread holding of SSNs among this population and (2) low participation in the types of programs that do not require SSNs.

For non-elderly adults, the results in most states are similar to what we see for the elderly. Reporting rates were 99 percent or higher in 31 states and above 95 percent in all but three states. Only California was below 90 percent, although the reporting rate in that state was less than half of what it was in most states at 49.1 percent. As we will show, this result for California reflects the high participation in restricted benefit programs.

For non-infant children the national reporting rate—96.2 percent—was markedly higher than the 88.0 percent observed for non-elderly adults, yet most states had lower SSN reporting rates for these children than for non-elderly adults. The reporting rate in California—84.8 percent—was not nearly as low as it was for non-elderly adults, so California did not depress the national results to quite the same degree that it did for non-elderly adults. One other state, Montana, fell below 90 percent, and two other states were below 95 percent. Of the remainder, 21 states had reporting rates of 99 percent or higher, and 26 states fell between 95 and 99 percent.

Table II.2. Percentage with an SSN in the Valid Range: Medicaid Enrollees by Age Group, Fourth Quarter FY 2009

State	Infants (< 1)	All Other Children (1 to 18)	Non-elderly Adults (19 to 64)	Elderly Adults (65+)	Total
United States	73.8	96.2	88.0	99.4	92.3
Alabama	59.1	98.2	99.1	100.0	97.0
Alaska	53.9	99.6	99.9	100.0	97.2
Arizona	90.9	95.0	91.2	99.7	93.4
Arkansas	84.0	99.7	98.9	100.0	98.8
California	65.5	84.8	49.1	98.2	68.2
Colorado	69.1	97.7	96.8	99.9	96.0
Connecticut	54.8	96.4	99.3	99.8	96.5
Delaware	47.5	94.4	95.3	99.9	93.5
District of Columbia	47.5	96.1	99.3	99.7	96.0
Florida	92.6	99.5	98.9	99.9	99.0
Georgia	71.9	96.3	98.5	99.7	95.7
Hawaii	87.4	98.6	100.0	100.0	98.9
Idaho	67.0	98.3	99.3	100.0	96.8
Illinois	91.9	98.9	99.7	99.9	99.0
Indiana	82.7	98.1	97.4	99.8	97.2
Iowa	80.8	99.7	99.7	100.0	98.9
Kansas	87.9	99.8	99.4	100.0	98.9
Kentucky	87.4	99.5	99.5	100.0	99.0
Louisiana	43.3	96.9	100.0	100.0	95.8
Maine	80.1	99.3	99.9	100.0	99.2
Maryland	55.5	97.3	98.8	99.9	96.2
Massachusetts	77.3	94.1	96.3	96.7	95.1
Michigan	86.5	96.7	98.6	99.8	97.3
Minnesota	87.0	99.3	99.3	99.5	98.8
Mississippi	57.6	99.9	99.9	100.0	97.6
Missouri	85.2	98.8	99.5	100.0	98.5
Montana	89.5	88.8	99.3	99.6	93.3
Nebraska	94.8	98.3	99.6	99.9	98.5
Nevada	41.4	96.8	98.9	100.0	93.9
New Hampshire	87.1	99.6	99.8	99.9	99.2
New Jersey	55.2	95.2	98.2	98.4	94.8
New Mexico	77.0	98.9	99.6	100.0	98.3
New York	53.7	96.2	97.6	99.7	94.6
North Carolina	82.1	99.1	98.9	100.0	98.3
North Dakota	97.0	99.9	100.0	100.0	99.8
Ohio	79.6	97.8	99.8	100.0	98.0
Oklahoma	81.4	98.1	98.5	100.0	97.5
Oregon	71.5	95.4	93.3	99.7	93.8
Pennsylvania	87.5	99.6	99.7	99.9	99.2
Rhode Island	91.4	99.0	99.9	98.6	99.1
South Carolina	60.4	99.5	99.7	100.0	97.6
South Dakota	78.7	99.5	99.4	99.9	98.4
Tennessee	87.3	99.9	99.8	100.0	99.4
Texas	69.9	98.9	97.3	98.7	96.6
Utah	93.1	99.5	98.6	99.8	98.7
Vermont	96.5	99.8	99.9	100.0	99.8
Virginia	49.1	98.5	99.4	100.0	96.4
Washington	72.1	98.1	99.8	99.9	97.7
West Virginia	89.5	99.8	100.0	100.0	99.5
Wisconsin	88.9	99.1	99.8	99.9	99.2
Wyoming	64.4	98.8	98.5	100.0	96.7

Source: State MSIS files, fourth quarter, FY 2009.

We found wide variation in the reporting of SSNs for infants. Five states were below 50 percent, with the lowest, Nevada, at 41.4 percent. Another seven states fell between 50 and 60 percent. Only eight states were above 90 percent, with the highest, North Dakota, at 97.0 percent. SSNs are not assigned to infants until a few weeks after birth, at the earliest. Those who are born with Medicaid coverage begin their period of enrollment without SSNs, which must be added to MSIS later. The earliest opportunity for a state to amend its MSIS data is at the next quarterly submission following the state's receipt of the infants' SSNs, but the updates are not always this timely. Even if SSNs are eventually added to the Medicaid records of nearly all infants, a cross-sectional snapshot will show many infants without them.

Restricted benefits are virtually nonexistent among infants, and they are rare among older children as well. Only California had as many as 10 percent of its older children enrolled with only restricted benefits, and most states had effectively none (Table II.3). Where restricted benefits factor into the prevalence of reported SSNs is among non-elderly adults. The incidence of restricted benefits is actually higher among the elderly than among non-elderly adults, but the very high reporting rates for SSNs among the elderly across all states imply that enrollees with restricted benefits have a minimal impact on the reporting of SSNs for this population—something that we will confirm below. Among non-elderly adults, enrollees with restricted benefits accounted for at least 10 percent of all enrollees in 26 states, and they exceeded 20 percent in six states. Enrollees with restricted benefits were fewer than 5 percent of all non-elderly adult enrollees in only 12 states.

We have seen at the national level that the reporting of SSNs in the valid range was lower for enrollees with restricted benefits than for enrollees with full benefits. For many states, there was little difference between the two enrollee populations, but for some states the differences

Table II.3. Percentage of Medicaid Enrollees with Restricted Benefits, by Age Group, Fourth Quarter FY 2009

State	Infants (< 1)	All Other Children (1 to 18)	Non-elderly Adults (19 to 64)	Elderly Adults (65+)	Total
United States	0.0	1.9	19.6	23.0	10.7
Alabama	0.0	0.1	40.4	55.4	20.9
Alaska	0.0	0.0	0.3	2.4	0.3
Arizona	0.1	3.8	12.0	26.6	8.8
Arkansas	0.0	0.8	44.4	36.5	16.8
California	0.3	10.9	52.2	3.5	29.8
Colorado	0.0	0.2	6.2	16.7	3.6
Connecticut	0.0	0.0	3.7	26.4	4.5
Delaware	0.0	2.2	17.1	52.2	12.8
District of Columbia	0.0	0.0	2.7	16.9	2.6
Florida	0.0	0.2	16.4	37.5	10.5
Georgia	0.0	0.0	12.7	44.4	8.0
Hawaii	0.0	0.0	1.1	8.7	1.2
Idaho	0.0	0.0	9.0	29.8	4.4
Illinois	0.0	0.0	5.5	11.2	2.8
Indiana	0.0	0.0	23.1	26.2	9.4
Iowa	0.0	1.8	14.7	21.4	8.5
Kansas	0.0	0.0	9.3	22.9	5.1
Kentucky	0.0	0.0	11.7	37.9	8.0
Louisiana	0.0	0.0	27.2	44.4	12.0
Maine	0.0	0.1	5.9	57.0	12.5
Maryland	0.1	1.0	12.5	29.4	7.3
Massachusetts	0.1	2.9	6.3	16.8	6.1
Michigan	0.0	0.9	9.9	11.8	5.1
Minnesota	0.0	1.8	8.4	11.2	5.3
Mississippi	0.0	0.3	24.1	44.1	13.4
Missouri	0.0	0.1	6.4	8.6	3.0
Montana	0.0	0.0	1.4	16.2	1.9
Nebraska	0.0	0.0	2.7	10.9	1.7
Nevada	0.0	0.1	13.3	40.6	7.8
New Hampshire	0.0	0.0	9.1	26.5	5.3
New Jersey	0.0	0.0	3.3	13.5	2.8
New Mexico	0.0	0.1	16.4	30.2	7.5
New York	0.0	0.7	3.2	13.6	3.4
North Carolina	0.0	0.1	15.3	21.1	7.3
North Dakota	0.0	0.0	5.6	24.1	4.7
Ohio	0.0	0.0	6.2	26.3	4.3
Oklahoma	0.0	0.1	16.1	16.6	6.0
Oregon	0.0	3.6	15.8	31.7	10.9
Pennsylvania	0.0	0.2	10.5	16.0	6.1
Rhode Island	0.0	0.0	2.6	13.5	2.8
South Carolina	0.0	0.5	18.3	12.9	7.4
South Dakota	0.0	0.0	9.9	33.3	5.6
Tennessee	0.0	0.0	4.9	36.2	5.4
Texas	0.0	0.1	12.0	34.4	6.3
Utah	0.0	0.2	2.3	8.5	1.4
Vermont	0.0	0.0	2.7	47.1	6.5
Virginia	0.0	0.0	12.4	27.3	6.5
Washington	0.0	1.1	17.7	19.4	7.6
West Virginia	0.0	0.0	10.2	37.6	7.8
Wisconsin	0.0	2.0	12.7	54.5	12.9
Wyoming	0.0	0.8	10.0	35.0	5.6

Source: State MSIS files, fourth quarter, FY 2009.

were stark, and these states drove the national results. Focusing on non-elderly adults, for whom the national reporting rate was only 41.5 percent, we found that in 20 states, SSNs in the valid range were reported for over 99 percent of the enrollees with restricted benefits, and in 12 other states the reporting rate was above 95 percent (Table II.4). In California, however, the reporting rate was only 3.1 percent, and in three other states the reporting rate was below 50 percent, leaving 15 states with SSN reporting rates between 50 and 95 percent. For the elderly, California reported SSNs in the valid range for only 57.9 percent of enrollees with restricted benefits, but only seven other states were below 99 percent, and all but one of these were above 97 percent.

Across all ages, six states reported SSNs for fewer than 75 percent of their enrollees with restricted benefits. California was the lowest by far at just 3.7 percent. Arizona was next lowest at 38.4 percent. Massachusetts (51.2 percent), Utah (62.0 percent), Oregon (63.8 percent), and Colorado (71.9 percent) rounded out this group. Three other states fell below 90 percent: Delaware (80.0 percent), Michigan (84.2 percent) and Wyoming (85.3 percent).

For enrollees with full benefits, the proportion with SSNs in the valid range was almost universally high across the states except among infants. Across all age groups the lowest reporting rate in any state was 93.1 percent (Montana), and only three other states were below 95 percent (Table II.5). For elderly adults, only three states were below 99.5 percent, and the lowest of these was 98.2 percent (New Jersey). For non-elderly adults, the incidence of near-perfect reporting was lower, but only four states were below 99 percent, the lowest being Indiana at 96.6 percent. Among non-infant children, rates below 99 percent were considerably more common, accounting for 27 states, but only two states were below 95 percent—California, narrowly, at 94.9 percent and Montana, much lower, at 88.8 percent. Among infants, the reporting rates were virtually the same as we saw in Table II.2 for all enrollees, as infants with restricted benefits

Table II.4. Percentage with an SSN in the Valid Range: Medicaid Enrollees with Restricted Benefits by Age Group, Fourth Quarter FY 2009

State	Infants (< 1)	All Other Children (1 to 18)	Non-elderly Adults (19 to 64)	Elderly Adults (65+)	Total
United States	43.4	13.7	41.5	98.1	51.1
Alabama	100.0	52.7	98.7	100.0	99.1
Alaska	--	--	100.0	100.0	100.0
Arizona	21.9	12.4	27.6	98.8	38.4
Arkansas	100.0	99.9	99.9	100.0	99.9
California	42.8	2.5	3.1	57.9	3.7
Colorado	0.0	2.5	52.2	99.6	71.9
Connecticut	--	--	99.9	99.9	99.9
Delaware	100.0	35.9	75.9	100.0	80.0
District of Columbia	--	0.0	84.5	99.3	92.7
Florida	100.0	81.4	96.9	99.8	98.2
Georgia	0.0	15.3	92.2	99.6	96.1
Hawaii	--	--	99.2	100.0	99.7
Idaho	--	--	100.0	100.0	100.0
Illinois	--	100.0	99.3	99.5	99.3
Indiana	--	100.0	99.9	100.0	99.9
Iowa	100.0	99.2	99.3	100.0	99.4
Kansas	--	6.6	94.0	99.9	96.3
Kentucky	--	15.0	98.6	100.0	99.3
Louisiana	--	95.7	99.9	100.0	100.0
Maine	0.0	18.2	98.8	100.0	99.6
Maryland	84.2	95.4	93.7	99.8	95.7
Massachusetts	17.3	15.3	46.6	80.9	51.2
Michigan	50.0	30.0	87.5	98.9	84.2
Minnesota	92.9	89.9	92.7	97.8	93.3
Mississippi	--	100.0	100.0	100.0	100.0
Missouri	--	100.0	99.9	100.0	99.9
Montana	--	--	99.1	99.6	99.5
Nebraska	--	--	99.9	100.0	99.9
Nevada	--	14.3	93.0	100.0	96.1
New Hampshire	--	--	100.0	100.0	100.0
New Jersey	100.0	13.7	77.3	99.7	91.0
New Mexico	--	49.5	98.0	99.9	98.2
New York	52.6	90.2	89.3	98.3	93.5
North Carolina	0.0	40.1	96.1	99.9	97.0
North Dakota	--	--	100.0	100.0	100.0
Ohio	--	--	100.0	100.0	100.0
Oklahoma	100.0	40.4	91.5	100.0	92.7
Oregon	50.0	15.1	58.1	99.1	63.8
Pennsylvania	100.0	88.8	98.1	99.6	98.4
Rhode Island	--	76.2	97.3	99.8	98.8
South Carolina	100.0	97.5	98.8	100.0	99.0
South Dakota	--	0.0	99.2	99.9	99.6
Tennessee	--	10.8	96.4	100.0	98.6
Texas	20.0	8.8	84.3	98.5	92.0
Utah	--	23.1	48.1	97.3	62.0
Vermont	--	--	100.0	100.0	100.0
Virginia	100.0	32.4	96.8	100.0	98.2
Washington	--	97.0	99.4	99.6	99.3
West Virginia	100.0	100.0	100.0	100.0	100.0
Wisconsin	100.0	97.5	98.8	99.9	99.3
Wyoming	0.0	9.9	85.2	99.9	85.3

Source: State MSIS files, fourth quarter, FY 2009.
-- No enrollees with restricted benefits.

Table II.5. Percentage with an SSN in the Valid Range: Medicaid Enrollees with Full Benefits by Age Group, Fourth Quarter FY 2009

State	Infants (< 1)	All Other Children (1 to 18)	Non-elderly Adults (19 to 64)	Elderly Adults (65+)	Total
United States	73.8	97.9	99.3	99.7	97.3
Alabama	59.1	98.2	99.3	100.0	96.4
Alaska	53.9	99.6	99.9	100.0	97.2
Arizona	91.0	98.3	99.9	100.0	98.8
Arkansas	84.0	99.7	98.0	99.9	98.6
California	65.6	94.9	99.3	99.6	95.6
Colorado	69.1	97.9	99.7	100.0	96.8
Connecticut	54.8	96.4	99.3	99.8	96.4
Delaware	47.4	95.7	99.3	99.9	95.4
District of Columbia	47.5	96.1	99.8	99.8	96.1
Florida	92.6	99.5	99.2	100.0	99.1
Georgia	71.9	96.3	99.5	99.8	95.6
Hawaii	87.4	98.6	100.0	100.0	98.9
Idaho	67.0	98.3	99.3	100.0	96.7
Illinois	91.9	98.9	99.7	100.0	98.9
Indiana	82.7	98.1	96.6	99.7	97.0
Iowa	80.8	99.7	99.8	100.0	98.8
Kansas	87.9	99.8	99.9	100.0	99.1
Kentucky	87.4	99.5	99.6	100.0	99.0
Louisiana	43.3	96.9	100.0	100.0	95.2
Maine	80.1	99.4	99.9	100.0	99.2
Maryland	55.5	97.4	99.5	99.9	96.3
Massachusetts	77.4	96.5	99.7	99.9	98.0
Michigan	86.5	97.2	99.8	100.0	98.0
Minnesota	87.0	99.5	99.9	99.8	99.1
Mississippi	57.6	99.9	99.8	100.0	97.3
Missouri	85.2	98.8	99.5	100.0	98.5
Montana	89.5	88.8	99.3	99.6	93.1
Nebraska	94.8	98.3	99.6	99.9	98.5
Nevada	41.4	96.8	99.8	100.0	93.7
New Hampshire	87.1	99.6	99.8	99.8	99.1
New Jersey	55.2	95.2	98.9	98.2	94.9
New Mexico	77.0	99.0	100.0	100.0	98.3
New York	53.7	96.2	97.9	99.9	94.6
North Carolina	82.1	99.1	99.4	100.0	98.4
North Dakota	97.0	99.9	100.0	100.0	99.8
Ohio	79.6	97.8	99.8	100.0	97.9
Oklahoma	81.4	98.2	99.9	100.0	97.9
Oregon	71.6	98.4	99.8	100.0	97.5
Pennsylvania	87.5	99.6	99.9	100.0	99.3
Rhode Island	91.4	99.0	99.9	98.4	99.1
South Carolina	60.4	99.5	99.9	100.0	97.5
South Dakota	78.7	99.5	99.4	99.8	98.3
Tennessee	87.3	99.9	100.0	100.0	99.4
Texas	69.9	99.0	99.1	98.8	96.9
Utah	93.1	99.6	99.8	100.0	99.2
Vermont	96.5	99.8	99.9	100.0	99.8
Virginia	49.1	98.5	99.8	100.0	96.3
Washington	72.1	98.1	99.9	100.0	97.5
West Virginia	89.5	99.8	100.0	100.0	99.4
Wisconsin	88.9	99.2	99.9	100.0	99.1
Wyoming	64.4	99.5	99.9	100.0	97.4

Source: State MSIS files, fourth quarter, FY 2009.

were exceedingly rare. While the lower incidence of reported SSNs among infants was not the

sole reason most states had SSNs for less than 99 percent of their enrollees with full benefits, the

reporting rate for infants was an important factor.

B. Validity of Reported SSNs: Fourth Quarter, FY 2009

The information collected from applications for SSNs is stored in an SSA database called

the Numident. Under an inter-agency agreement with SSA, the Census Bureau obtains regular

updates to the Numident and has enhanced this database with information collected from its

censuses, surveys, and other administrative records.[9] These enhancements increase the Census

Bureau's ability to link its Numident data to other databases, which plays a critical role in the

Census Bureau's administrative record research. To the extent that these enhancements have

improved the accuracy and completeness of the gender and date of birth information in the

Census Bureau's Numident, we can expect a more thorough validation of the MSIS SSNs than

SSA would be able to provide. Section 1 presents the results of our validation of SSNs from the

fourth quarter of FY 2009, and Section 2 examines why SSNs passed or failed validation.

1.   Results of SSN Validation

The Census Bureau validation program applies four tests that involve comparing gender and

the three components of the date of birth (year, month, and day) between the MSIS record and

the Numident.

- Test 1: the year of birth differs by no more than one year, and either the month or day
  agrees exactly, and gender agrees
- Test 2: the year of birth agrees, and the month and day are inverted

---

[9] One consequence of these enhancements is that the Census Bureau Numident falls under Title 13, the law authorizing the Census Bureau to collect census data and requiring the Bureau to protect its confidentiality. An important implication of the Title 13 protection is that the Census Bureau cannot share identifiable data from its Numident with other agencies or organizations. While Mathematica as a CMS contractor was able to use the Census Bureau Numident in a protected environment at the Census Bureau to estimate validation rates for subpopulations, Mathematica could not remove from the Census Bureau the validation results for individual enrollees.

- Test 3: the month and day of birth agree, and gender agrees
- Test 4: the year, month, and day of birth agree

An SSN is classified as valid if the MSIS record on which it appears passes any one of these tests.

The four tests can be performed only when the MSIS SSN is found in the Numident. Therefore, there are three possible outcomes of the Census Bureau validation procedure: (1) an SSN passes validation, (2) an SSN is found in the Numident but fails validation, and (3) an SSN is not found in the Numident. We applied the validation tests to SSNs that fell within the valid range, as defined earlier.

a. National Results

Of the 50.8 million records submitted for validation, 99.37 percent passed (Table II.6). Another 0.53 percent of the records had SSNs and were found in the Numident but failed all four of the tests. The SSNs on the remaining 0.10 percent of the records were not in the Numident.

Table II.6. Outcome of Census Bureau Validation Tests: Medicaid Enrollees with an SSN in the Valid Range, by Age, Fourth Quarter, FY 2009

Outcome of Validation Test	Infants (< 1)	All Other Children (1 to 18)	Non-elderly Adult (19 to 64)	Elderly Adult (65+)	Total
All Enrollees with SSNs in the Valid Range					
Total	1,728,116	25,076,076	18,617,639	5,366,106	50,787,937
SSN passes Census Bureau validation	1,710,802	24,894,133	18,546,256	5,317,846	50,469,037
SSN is in the Numident but fails validation	13,781	143,795	65,096	47,244	269,916
SSN is not in the Numident	3,533	38,148	6,287	1,016	48,984
Percentage Distribution					
Total	100.00	100.00	100.00	100.00	100.00
SSN passes Census Bureau validation	99.00	99.27	99.62	99.10	99.37
SSN is in the Numident but fails validation	0.80	0.57	0.35	0.88	0.53
SSN is not in the Numident	0.20	0.15	0.03	0.02	0.10

Source: State MSIS files, fourth quarter, FY 2009.

It is apparent from the national results that validation failures were not a serious issue in any age group. Even among infants, whose SSNs we had expected to be the most problematic, 99.00 percent of those we tested passed validation. This compared to 99.27 percent for older children, 99.62 percent for non-elderly adults, and 99.10 percent for elderly adults.[10]

In every age group, most of the SSNs that failed validation were found in the Numident but with different demographic characteristics than what was recorded in the MSIS records. For infants, 0.80 percent of the SSNs failed to match demographic characteristics whereas 0.20 percent were not in the Numident. For non-infant children, 0.57 percent failed to match demographic characteristics whereas 0.15 percent were not in the Numident. For non-elderly adults, 0.35 percent failed to match demographic characteristics whereas 0.03 percent were not in the Numident. For elderly adults, 0.88 percent failed to match demographic characteristics whereas 0.02 percent were not in the Numident.

That the validation rate for elderly adult SSNs should be lower than that for non-infant children and non-elderly adults, if only slightly, is surprising because MSIS records for the elderly were the most likely to have SSNs in the valid range. It could be that elderly Medicaid applicants were more likely than younger applicants to misreport a date of birth. It is also possible that an elderly applicant who is a dual eligible is more likely than another applicant to report a spouse's SSN as his or her own, as this may be the SSN used in that individual's HIC. The comparatively high rate of elderly SSNs that are in the Numident but fail validation is consistent with both of these explanations.

---

[10] We did not estimate validation rates separately for enrollees with full versus restricted benefits, as range checks suggested that, despite the substantial difference in reporting rates, the SSNs reported for these two groups were likely to be of comparable quality. For enrollees with full benefits, 97 percent reported an SSN, and only 0.03 percent of the reported SSNs fell outside the valid range. For enrollees with restricted benefits, only 51 percent reported an SSN, yet only 0.02 percent of the reported SSNs fell outside the valid range.

b.   State Results

Table II.7 reports for each state the proportion of SSNs, by age group, that passed validation. In every age group, the validation rates were concentrated in a very narrow range. For infants, three states had validation rates of 99.9 percent, and only three states had rates below 98 percent. Of these, New Jersey was lowest at 96.0 percent, followed by the District of Columbia at 96.4 percent and Connecticut at 97.5 percent. For older children, five states shared the top validation rate of 99.9 percent, and less than a quarter of the states had rates below 99 percent. New Jersey was lowest at 97.6 percent, but no other state was below 98 percent. For non-elderly adults, 11 states had validation rates of 99.9 percent, and only two states were below 99 percent. New Jersey was again the lowest of these at 98.4 percent, followed by Delaware at 98.6 percent. For elderly adults, only one state was as high as 99.8 percent, but only one state was below 98 percent: the District of Columbia had a validation rate of 97.7 percent. There is no evident pattern by region or state size.

The overall validation rates for all age groups combined fell within a range of only 2.0 percentage points, from 97.9 percent for New Jersey to 99.9 percent for four states. Only seven states were below 99 percent.

In discussing our results at the national level, we noted that most of the validation failures were due to SSNs that were found in the Numident but with different demographic characteristics than those recorded in the MSIS records. This general pattern was repeated in virtually every state (compare Tables II.8 and II.9). For all age groups combined there were only two states—Arizona and Florida—in which the more common reason for failure was an SSN not being in the Numident. This was also true among non-infant children—and for the same two states. For the other three age groups, there was no state in which SSNs not in the Numident

Table II.7. Percentage of SSNs Passing Census Bureau Validation: Medicaid Enrollees with an SSN in the Valid Range by Age and State, Fourth Quarter, FY 2009

State	Infants (< 1)	All Other Children (1 to 18)	Non-elderly Adults (19 to 64)	Elderly Adults (65+)	Total
United States	99.0	99.3	99.6	99.1	99.4
Alabama	98.9	99.2	99.8	98.8	99.3
Alaska	98.2	99.2	99.8	98.5	99.3
Arizona	99.7	99.0	99.6	98.9	99.3
Arkansas	99.7	99.8	99.8	99.3	99.8
California	99.1	99.7	99.4	99.2	99.5
Colorado	98.6	98.8	99.6	99.2	99.1
Connecticut	97.5	98.3	99.3	98.8	98.8
Delaware	98.2	98.4	98.6	98.2	98.5
District of Columbia	96.4	98.1	99.1	97.7	98.5
Florida	99.1	98.9	99.7	99.2	99.2
Georgia	98.3	98.4	99.5	98.5	98.7
Hawaii	99.5	99.8	99.9	99.3	99.8
Idaho	99.9	99.9	99.9	99.8	99.9
Illinois	99.4	98.4	99.6	98.9	98.9
Indiana	99.5	99.9	99.8	99.6	99.8
Iowa	99.2	99.1	99.7	99.6	99.4
Kansas	99.6	99.9	99.9	99.6	99.8
Kentucky	99.8	99.7	99.9	99.3	99.7
Louisiana	98.2	99.0	99.8	99.2	99.2
Maine	99.1	98.9	99.7	99.5	99.4
Maryland	98.1	99.0	99.5	98.7	99.1
Massachusetts	98.9	99.5	99.5	99.0	99.4
Michigan	99.5	99.7	99.8	99.4	99.7
Minnesota	99.5	99.7	99.7	99.4	99.7
Mississippi	99.1	99.7	99.7	98.7	99.5
Missouri	99.8	99.9	99.9	99.6	99.9
Montana	99.9	99.8	99.8	99.7	99.8
Nebraska	99.9	99.8	99.9	99.7	99.9
Nevada	99.1	99.4	99.8	99.5	99.5
New Hampshire	98.7	99.3	99.7	99.3	99.4
New Jersey	96.0	97.6	98.4	98.3	97.9
New Mexico	98.4	99.3	99.7	98.8	99.4
New York	98.7	99.5	99.5	98.9	99.4
North Carolina	99.7	99.7	99.7	99.2	99.7
North Dakota	99.7	99.9	99.9	99.7	99.9
Ohio	98.8	99.0	99.7	99.1	99.3
Oklahoma	99.5	99.5	99.6	99.5	99.5
Oregon	98.4	98.7	99.4	99.6	99.1
Pennsylvania	99.1	99.2	99.7	99.0	99.4
Rhode Island	99.4	99.8	99.8	98.4	99.6
South Carolina	98.5	99.2	99.7	98.0	99.2
South Dakota	98.4	98.9	99.8	99.7	99.2
Tennessee	99.5	99.7	99.9	99.3	99.7
Texas	98.2	99.1	99.5	99.0	99.2
Utah	99.8	99.8	99.7	99.5	99.8
Vermont	99.8	99.8	99.7	99.6	99.7
Virginia	98.6	99.4	99.7	99.1	99.4
Washington	99.0	99.8	99.9	99.6	99.8
West Virginia	99.7	99.8	99.9	99.6	99.8
Wisconsin	99.7	99.8	99.9	99.7	99.8
Wyoming	98.3	98.6	99.3	99.5	98.8

Source: State MSIS files, fourth quarter, FY 2009.

Table II.8. Percentage of SSNs Not Found in the Numident: Medicaid Enrollees with an SSN in the Valid Range by Age and State, Fourth Quarter, FY 2009

State	Infants (< 1)	All Other Children (1 to 18)	Non-elderly Adults (19 to 64)	Elderly Adults (65+)	Total
United States	0.2	0.2	0.0	0.0	0.1
Alabama	0.1	0.1	0.0	0.0	0.1
Alaska	0.2	0.1	0.0	0.0	0.1
Arizona	0.1	0.9	0.1	0.1	0.4
Arkansas	0.1	0.0	0.0	0.0	0.0
California	0.1	0.0	0.1	0.0	0.0
Colorado	0.4	0.3	0.0	0.0	0.2
Connecticut	0.5	0.2	0.1	0.0	0.1
Delaware	0.1	0.3	0.2	0.0	0.2
District of Columbia	0.0	0.0	0.0	0.0	0.0
Florida	0.3	0.9	0.1	0.1	0.5
Georgia	0.7	0.4	0.0	0.0	0.3
Hawaii	0.2	0.0	0.0	0.0	0.0
Idaho	0.0	0.0	0.0	0.0	0.0
Illinois	0.1	0.0	0.0	0.0	0.0
Indiana	0.0	0.0	0.0	0.0	0.0
Iowa	0.1	0.1	0.0	0.0	0.1
Kansas	0.1	0.0	0.0	0.0	0.0
Kentucky	0.0	0.0	0.0	0.0	0.0
Louisiana	0.5	0.2	0.0	0.0	0.1
Maine	0.3	0.1	0.0	0.0	0.1
Maryland	0.2	0.1	0.0	0.0	0.1
Massachusetts	0.2	0.1	0.1	0.0	0.1
Michigan	0.2	0.1	0.0	0.0	0.0
Minnesota	0.1	0.0	0.0	0.0	0.0
Mississippi	0.3	0.0	0.0	0.0	0.0
Missouri	0.1	0.0	0.0	0.0	0.0
Montana	0.0	0.0	0.0	0.0	0.0
Nebraska	0.0	0.0	0.0	0.0	0.0
Nevada	0.0	0.1	0.0	0.0	0.1
New Hampshire	0.1	0.1	0.0	0.0	0.1
New Jersey	0.7	0.3	0.2	0.0	0.2
New Mexico	0.0	0.0	0.0	0.0	0.0
New York	0.1	0.1	0.0	0.0	0.0
North Carolina	0.1	0.0	0.0	0.0	0.0
North Dakota	0.1	0.0	0.0	0.0	0.0
Ohio	0.2	0.1	0.0	0.0	0.1
Oklahoma	0.1	0.1	0.0	0.0	0.1
Oregon	0.0	0.0	0.0	0.0	0.0
Pennsylvania	0.1	0.1	0.0	0.0	0.1
Rhode Island	0.2	0.0	0.0	0.1	0.0
South Carolina	0.3	0.2	0.0	0.0	0.1
South Dakota	0.1	0.1	0.0	0.0	0.1
Tennessee	0.0	0.0	0.0	0.0	0.0
Texas	0.4	0.2	0.0	0.0	0.1
Utah	0.1	0.1	0.1	0.0	0.1
Vermont	0.0	0.0	0.0	0.0	0.0
Virginia	0.3	0.1	0.0	0.0	0.1
Washington	0.1	0.0	0.0	0.0	0.0
West Virginia	0.0	0.0	0.0	0.0	0.0
Wisconsin	0.1	0.0	0.0	0.0	0.0
Wyoming	0.1	0.2	0.1	0.0	0.1

Source: State MSIS files, fourth quarter, FY 2009.

Table II.9. Percentage of SSNs in the Numident but Failing Census Bureau Validation: Medicaid Enrollees with an SSN in the Valid Range by Age and State, Fourth Quarter, FY 2009

State	Infants (< 1)	All Other Children (1 to 18)	Non-elderly Adults (19 to 64)	Elderly Adults (65+)	Total
United States	0.8	0.6	0.3	0.9	0.5
Alabama	1.0	0.7	0.2	1.2	0.6
Alaska	1.6	0.7	0.2	1.5	0.6
Arizona	0.2	0.1	0.3	1.0	0.3
Arkansas	0.2	0.2	0.2	0.7	0.2
California	0.7	0.3	0.5	0.7	0.4
Colorado	1.0	0.9	0.4	0.8	0.8
Connecticut	2.0	1.5	0.6	1.2	1.1
Delaware	1.7	1.4	1.1	1.8	1.3
District of Columbia	3.6	1.9	0.8	2.2	1.5
Florida	0.6	0.3	0.2	0.7	0.3
Georgia	1.0	1.2	0.4	1.5	1.0
Hawaii	0.3	0.2	0.1	0.7	0.2
Idaho	0.1	0.1	0.1	0.2	0.1
Illinois	0.5	1.6	0.4	1.1	1.1
Indiana	0.5	0.1	0.2	0.4	0.2
Iowa	0.7	0.8	0.3	0.4	0.5
Kansas	0.3	0.1	0.1	0.4	0.1
Kentucky	0.2	0.3	0.1	0.7	0.3
Louisiana	1.3	0.8	0.2	0.8	0.6
Maine	0.6	1.0	0.2	0.5	0.6
Maryland	1.8	0.9	0.5	1.3	0.8
Massachusetts	0.9	0.4	0.5	1.0	0.5
Michigan	0.3	0.2	0.2	0.6	0.2
Minnesota	0.4	0.3	0.3	0.6	0.3
Mississippi	0.6	0.3	0.3	1.3	0.4
Missouri	0.1	0.1	0.1	0.4	0.1
Montana	0.1	0.1	0.2	0.3	0.2
Nebraska	0.1	0.1	0.1	0.3	0.1
Nevada	0.9	0.5	0.2	0.5	0.4
New Hampshire	1.2	0.6	0.3	0.7	0.5
New Jersey	3.2	2.1	1.4	1.6	1.8
New Mexico	1.6	0.7	0.3	1.2	0.6
New York	1.2	0.5	0.5	1.1	0.6
North Carolina	0.2	0.2	0.2	0.8	0.3
North Dakota	0.3	0.1	0.1	0.3	0.1
Ohio	1.0	0.9	0.2	0.9	0.6
Oklahoma	0.4	0.5	0.4	0.5	0.4
Oregon	1.6	1.3	0.6	0.4	0.9
Pennsylvania	0.8	0.7	0.3	1.0	0.5
Rhode Island	0.4	0.1	0.2	1.6	0.4
South Carolina	1.2	0.6	0.3	2.0	0.7
South Dakota	1.5	1.0	0.2	0.3	0.8
Tennessee	0.5	0.2	0.1	0.7	0.3
Texas	1.4	0.7	0.4	1.0	0.7
Utah	0.1	0.1	0.2	0.5	0.2
Vermont	0.2	0.1	0.3	0.4	0.2
Virginia	1.1	0.5	0.3	0.9	0.5
Washington	0.9	0.2	0.1	0.4	0.2
West Virginia	0.3	0.2	0.1	0.4	0.2
Wisconsin	0.2	0.2	0.1	0.3	0.2
Wyoming	1.6	1.2	0.6	0.5	1.0

Source: State MSIS files, fourth quarter, FY 2009.

were more common than SSNs failing to match on demographic characteristics. The two reasons for SSN validation failure were equally common among infants in Missouri and Utah and among non-infant children in Utah.

We can expand our results to express the number of valid SSNs as a fraction of *all* MSIS records, not just those with reported SSNs in the valid range.[11] This is done in Table II.10, which shows that 91.7 percent of all MSIS records had valid (or validated) SSNs. This figure was lowest for infants, at 73.1 percent. For all other children, it was 95.5 percent, and for the elderly it was 98.5 percent, but for non-elderly adults, this fraction was a good deal lower at 87.6 percent. The low rate for non-elderly adults is due primarily to California, in which only 48.8 percent of the MSIS records for this age group had valid SSNs. No other state had a rate below 90 percent among non-elderly adults, and only three states had rates below 95 percent (Arizona, Delaware, and Oregon). In nearly every other state, the fraction of MSIS records with valid SSNs was similar between elderly and non-elderly adults.

Another way to express the incidence of valid SSNs is as a fraction of all reported SSNs. This is, if an MSIS record included an SSN, how often was it valid? Table II.11 includes these additional rates, which are virtually identical to what we reported in Table II.7. The close resemblance between the two tables occurs because there were very few reported SSNs that fell outside the valid range, as we defined it earlier—only a little more than 13,000 nationally, as reported in Table II.1. We have included these results because validation rates for all reported

---

[11] As explained earlier, our estimates exclude the MSIS records of children who were enrolled only in S-CHIP during the quarter.

Table II.10. Percentage of All MSIS Records with a Validated SSN: Medicaid Enrollees by Age and State, Fourth Quarter, FY 2009

State	Infants (< 1)	All Other Children (1 to 18)	Non-elderly Adults (19 to 64)	Elderly Adults (65+)	Total
United States	73.1	95.5	87.6	98.5	91.7
Alabama	58.4	97.4	98.8	98.8	96.3
Alaska	53.0	98.8	99.7	98.5	96.5
Arizona	90.7	94.0	90.9	98.6	92.8
Arkansas	83.7	99.5	98.7	99.3	98.6
California	65.0	84.5	48.8	97.4	67.9
Colorado	68.1	96.5	96.4	99.1	95.1
Connecticut	53.4	94.8	98.7	98.6	95.4
Delaware	46.6	92.9	94.0	98.1	92.0
District of Columbia	45.8	94.3	98.5	97.5	94.5
Florida	91.7	98.3	98.6	99.1	98.2
Georgia	70.7	94.7	98.1	98.2	94.4
Hawaii	87.0	98.4	99.8	99.3	98.7
Idaho	66.9	98.2	99.3	99.7	96.7
Illinois	91.4	97.3	99.3	98.8	97.9
Indiana	82.4	98.0	97.2	99.4	97.1
Iowa	80.1	98.8	99.4	99.6	98.3
Kansas	87.6	99.6	99.3	99.6	98.8
Kentucky	87.3	99.2	99.4	99.3	98.8
Louisiana	42.5	95.8	99.8	99.1	95.1
Maine	79.3	98.2	99.6	99.5	98.6
Maryland	54.4	96.3	98.3	98.6	95.4
Massachusetts	76.5	93.6	95.8	95.7	94.6
Michigan	86.0	96.4	98.4	99.3	97.0
Minnesota	86.6	99.0	99.0	98.9	98.5
Mississippi	57.1	99.6	99.6	98.7	97.2
Missouri	85.0	98.7	99.4	99.5	98.4
Montana	89.4	88.6	99.1	99.3	93.1
Nebraska	94.7	98.1	99.5	99.7	98.4
Nevada	41.1	96.2	98.7	99.5	93.4
New Hampshire	85.9	98.8	99.6	99.2	98.5
New Jersey	53.0	92.9	96.6	96.8	92.8
New Mexico	75.8	98.2	99.3	98.8	97.7
New York	53.0	95.7	97.1	98.6	94.0
North Carolina	81.9	98.8	98.7	99.1	97.9
North Dakota	96.7	99.8	99.9	99.7	99.7
Ohio	78.6	96.9	99.5	99.0	97.3
Oklahoma	81.0	97.6	98.1	99.5	97.1
Oregon	70.4	94.2	92.7	99.3	93.0
Pennsylvania	86.7	98.8	99.4	98.9	98.6
Rhode Island	90.9	98.8	99.6	97.0	98.7
South Carolina	59.5	98.7	99.4	98.0	96.9
South Dakota	77.4	98.3	99.2	99.6	97.6
Tennessee	86.9	99.7	99.7	99.3	99.1
Texas	68.6	98.1	96.9	97.7	95.8
Utah	92.9	99.3	98.3	99.3	98.4
Vermont	96.4	99.6	99.6	99.6	99.5
Virginia	48.4	97.9	99.1	99.0	95.9
Washington	71.4	97.8	99.7	99.5	97.4
West Virginia	89.2	99.6	99.9	99.6	99.3
Wisconsin	88.4	98.9	99.6	99.7	99.0
Wyoming	63.3	97.4	97.8	99.5	95.6

Source: State MSIS files, fourth quarter, FY 2009.

Table II.11. Percentage of Reported SSNs Validated: Medicaid Enrollees by Age and State, Fourth Quarter, FY 2009

State	Infants (< 1)	All Other Children (1 to 18)	Non-elderly Adults (19 to 64)	Elderly Adults (65+)	Total
United States	98.9	99.2	99.6	99.1	99.3
Alabama	98.9	99.2	99.8	98.8	99.3
Alaska	98.2	99.1	99.8	98.5	99.3
Arizona	99.7	99.0	99.6	98.9	99.3
Arkansas	99.7	99.8	99.8	99.3	99.8
California	99.1	99.7	99.4	99.2	99.5
Colorado	98.6	98.8	99.5	99.2	99.0
Connecticut	97.5	98.3	99.3	98.8	98.7
Delaware	98.2	98.4	98.6	98.2	98.5
District of Columbia	96.4	98.1	99.1	97.7	98.5
Florida	98.9	98.8	99.7	99.2	99.2
Georgia	98.3	98.3	99.5	98.5	98.7
Hawaii	99.5	99.8	99.9	99.3	99.8
Idaho	99.9	99.9	99.9	99.8	99.9
Illinois	99.4	98.4	99.6	98.9	98.9
Indiana	99.5	99.9	99.8	99.6	99.8
Iowa	99.2	99.1	99.7	99.6	99.4
Kansas	99.6	99.9	99.9	99.6	99.8
Kentucky	99.8	99.7	99.8	99.3	99.7
Louisiana	98.2	99.0	99.8	99.2	99.2
Maine	99.1	98.8	99.7	99.5	99.4
Maryland	98.0	99.0	99.5	98.7	99.1
Massachusetts	98.9	99.5	99.4	99.0	99.4
Michigan	99.5	99.7	99.8	99.4	99.7
Minnesota	99.5	99.7	99.7	99.4	99.7
Mississippi	99.0	99.7	99.7	98.7	99.5
Missouri	99.8	99.9	99.9	99.6	99.9
Montana	96.7	89.1	99.6	99.7	94.0
Nebraska	99.8	98.8	99.6	99.7	99.1
Nevada	99.1	99.4	99.8	99.5	99.5
New Hampshire	98.6	99.2	99.7	99.3	99.4
New Jersey	95.7	97.5	98.4	98.3	97.9
New Mexico	98.3	99.3	99.6	98.8	99.3
New York	98.6	99.4	99.5	98.9	99.4
North Carolina	99.6	99.7	99.7	99.2	99.7
North Dakota	99.7	99.9	99.9	99.7	99.9
Ohio	98.8	99.0	99.7	99.1	99.3
Oklahoma	99.5	99.5	99.5	99.5	99.5
Oregon	98.4	98.7	99.4	99.6	99.1
Pennsylvania	99.1	99.2	99.7	99.0	99.4
Rhode Island	99.4	99.8	99.8	98.3	99.6
South Carolina	98.4	99.2	99.7	98.0	99.2
South Dakota	98.4	98.9	99.8	99.7	99.2
Tennessee	99.5	99.7	99.9	99.3	99.7
Texas	98.2	99.1	99.5	99.0	99.2
Utah	99.8	99.8	99.6	99.5	99.7
Vermont	99.8	99.8	99.7	99.6	99.7
Virginia	98.2	99.4	99.7	99.0	99.4
Washington	99.0	99.8	99.9	99.6	99.8
West Virginia	99.6	99.8	99.9	99.6	99.8
Wisconsin	98.6	99.8	99.9	99.7	99.8
Wyoming	98.3	98.5	99.2	99.5	98.8

Source: State MSIS files, fourth quarter, FY 2009.

SSNs may be of broader interest than validation rates for the large subset of reported SSNs that met basic range checks.[12,13]

Across all age groups the largest difference between the two rates occurred in Montana, where 94.0 percent of all reported SSNs were found to be valid compared to 99.8 percent of the SSNs that fell into the valid range (compare Tables II.11 and II.7). This exceptional result in Montana can be attributed to the fact that the state accounted for 41 percent of the 13,000 SSNs nationally that fell outside the valid range.[14] The next largest difference was 0.8 percentage points in Nebraska. Only four other states had differences as large as 0.1 percentage points.

c. Numident and the High Group Test

The high group test, discussed in Chapter I, is used by CMS and its contractors to identify SSNs that were never issued and, therefore, must be incorrect. To assess the effectiveness of the high group test, which is more rigorous than the basic range checks we applied above, we compared the number of fourth quarter 2009 MSIS SSNs failing the high group test and any other range checks, by state, with the number of SSNs that could not be matched to the Numident.[15] Overall, the number of SSNs that failed the high group test was 69.7 percent of the number of SSNs that

---

[12] The denominators in Tables II.6, II.7, and II.8 exclude 28 reported SSNs with missing MSIS dates of birth because we could not classify such records by age and, therefore, did not include them in our Census Bureau tabulations. For completeness, however, the final column in Tables II.9 and II.10 includes these SSNs (although their impact is negligible). All these SSNs would fail the Census Bureau's validation tests because of their missing dates of birth, but, in fact, their validity cannot be determined.

[13] The Wisconsin MSIS file at CMS includes 1,254 more records with SSNs in the valid range than the MSIS file at the Census Bureau. Wisconsin's MSIS data for the final quarters of FY 2009 were still awaiting approval at CMS. We speculate that the Census Bureau may have received a preliminary file. We have removed these records—by age group—from the denominators in Tables II.9 and II.10.

[14] In Montana, 5.9 percent of the reported SSNs were outside the valid range; specifically, they were high. The next largest fraction of reported SSNs outside the valid range was 0.7 percent in Nebraska. No other state exceeded 0.1 percent on this measure.

[15] The high group test was run separately from the identification of MSIS SSNs outside the valid range, described earlier, and did not play a role in determining what SSNs we attempted to match to the Numident. For the comparison reported here, however, the SSNs that we defined as outside the valid range were counted as failing the high group test and not in the Numident.

could not be found in the Numident (Table II.12). In other words, the number of SSNs that failed the high group test was more than two-thirds of the number of SSNs not found in the Numident.

The results differed by age group. For infants the number of SSNs failing the high group test was 117.6 percent of the number not in the Numident, and in virtually every state the number of SSNs failing the high group test was at least as high as the number not in the Numident. More specifically, in 15 states the number failing the high group test and the number not in the Numident were essentially identical (ratios of 100.0 percent), and in 13 other states the number failing the high group test was within 10 percent of the number not in the Numident. Among all other children, the number of SSNs failing the high group test was only 61.4 percent of the number not in the Numident—about half the ratio for infants. The ratio among non-elderly adults was 92.7 percent whereas the ratio among elderly adults was only 55.2 percent. For both of the adult groups there were many states with ratios of 100.0 percent or very close to that value—even among the elderly, where the national ratio was brought down by a relatively small number of states with low but nonzero ratios. California and Florida accounted for more than half of the SSNs that were not in the Numident, but only 24 percent of these SSNs failed the high group test (calculated from data not shown). Most of the states with ratios of 0.0 percent—and some of those with ratios of 100 percent—had only one SSN not in the Numident.

Discrepancies between the high group test and whether SSNs are actually found in the Numident arise from two causes. The high group test is based on the group code that SSA reports as being used with each area code in a specific month. The Numident used by the Census Bureau has a less explicit reference period. It would appear, for example, that the 2009

Table II.12. Percentage of Reported SSNs Not in the Numident That Failed the High Group Test: Medicaid Enrollees by Age and State, Fourth Quarter, FY 2009

State	Infants (< 1)	All Other Children (1 to 18)	Non-elderly Adults (19 to 64)	Elderly Adults (65+)	Total
United States	117.6	61.4	92.7	55.2	69.7
Alabama	100.0	79.3	100.0	100.0	81.4
Alaska	60.0	79.3	100.0	0.0	78.8
Arizona	257.8	4.4	70.3	23.9	10.9
Arkansas	100.0	50.0	107.7	0.0	67.0
California	100.9	84.1	100.3	65.2	91.9
Colorado	100.0	95.0	100.8	90.0	96.0
Connecticut	100.0	92.3	100.6	100.0	95.1
Delaware	100.0	96.1	100.0	100.0	98.2
District of Columbia	n.a.	0.0	0.0	0.0	0.0
Florida	103.1	11.4	35.6	12.6	17.2
Georgia	125.1	95.5	98.6	86.2	98.8
Hawaii	94.4	75.9	75.0	33.3	79.6
Idaho	n.a.	10.5	n.a.	n.a.	10.5
Illinois	442.9	36.2	58.3	0.0	141.2
Indiana	100.0	4.6	71.4	0.0	12.2
Iowa	100.0	91.0	96.7	100.0	92.2
Kansas	100.0	63.9	100.0	100.0	75.5
Kentucky	433.3	63.6	103.3	n.a.	80.4
Louisiana	100.0	96.7	83.3	50.0	96.8
Maine	94.7	91.5	96.4	100.0	93.4
Maryland	102.3	96.4	98.6	90.9	97.2
Massachusetts	113.8	87.0	97.4	82.8	93.8
Michigan	226.3	96.2	111.6	100.0	119.3
Minnesota	134.4	91.1	102.1	100.0	100.8
Mississippi	111.8	99.5	106.7	100.0	103.1
Missouri	193.3	73.7	102.0	100.0	90.8
Montana	101.2	99.9	100.0	n.a.	99.9
Nebraska	100.0	99.0	99.5	100.0	99.1
Nevada	120.0	82.5	66.7	0.0	82.6
New Hampshire	100.0	87.5	90.0	100.0	89.0
New Jersey	101.1	98.4	99.7	95.3	98.8
New Mexico	166.7	51.5	98.5	100.0	80.1
New York	100.0	81.5	94.5	76.5	88.8
North Carolina	101.3	62.9	99.2	33.3	78.8
North Dakota	0.0	63.6	n.a.	n.a.	53.8
Ohio	108.3	93.5	102.0	97.4	95.9
Oklahoma	303.4	88.4	100.8	100.0	104.8
Oregon	175.0	53.8	216.7	n.a.	83.7
Pennsylvania	114.6	92.6	97.2	91.3	94.7
Rhode Island	60.0	68.4	86.7	100.0	84.0
South Carolina	100.0	89.9	100.0	100.0	91.6
South Dakota	100.0	90.0	100.0	n.a.	90.7
Tennessee	138.5	30.6	90.0	0.0	52.2
Texas	100.9	96.6	101.3	91.7	97.5
Utah	259.1	99.1	101.0	100.0	110.8
Vermont	n.a.	48.0	100.0	100.0	78.0
Virginia	101.3	91.6	99.0	100.0	94.2
Washington	106.7	66.9	88.9	n.a.	73.8
West Virginia	183.3	73.4	100.0	n.a.	84.0
Wisconsin	107.2	92.0	101.4	81.8	101.0
Wyoming	100.0	94.5	100.0	100.0	95.5

Source: State MSIS files, fourth quarter, FY 2009.

Numident does not include all SSNs issued in 2009.[16] If the Numident postdates the parameters used in the high group test, the Numident will include SSNs with group numbers that for some area codes were not yet in use and, therefore, will be deemed invalid by the high group test. Because the area codes are geographic, this pattern will vary by state, and so will the magnitudes of the discrepancies between the high group test and the Numident.

When the Numident does not include SSNs with group codes beyond those used in the high group test, the high group test will tend to understate the number of SSNs not in the Numident. This is because SSNs in the group that is currently being assigned to an area will not fail the high group test, but not all of the possible SSNs in that group will have been assigned. This is relevant only to new SSNs, however, so it will have an age pattern. New SSNs are a very small fraction of all SSNs among adults and older children, but all infant SSNs are less than a year old, and there may be a lot of new SSNs among young children as well.

We see that in some states the number of infant SSNs failing the high group test was two to four times the number not in the Numident. In Illinois, for example, the ratio was 442.9 percent, and the actual counts on which this ratio was based were substantially larger than in most states: 248 failing the high group test versus 56 not in the Numident. We also note that in Illinois the ratio among non-infant children was much smaller at 36.2 percent, but we do not see this pattern consistently in other states. In Arizona and Kentucky there were high ratios among infants and low ratios among non-infant children, but in Michigan, Oklahoma, and Utah there were ratios well above 100 percent for infants and yet the ratios for non-infant children are not far below 100

---

[16] Initially, we validated the SSNs in the fourth quarter 2009 MSIS data against a 2009 Numident, but it was clear from our results that the MSIS data contained valid SSNs that were not included in the 2009 Numident. When we validated the 2009 MSIS data against a 2010 Numident, 32.90 percent of the infant SSNs in the valid range passed validation whereas they had not done so with the 2009 Numident. For older children this fraction was only 0.15 percent, which compared to 0.20 percent for non-elderly adults and 0.19 percent for elderly adults. These latter rates, while small, show that, among the Medicaid-eligible population, SSNs continue to be issued to persons 65 and older—and at rates comparable to all younger persons except infants.

percent. We can understand why the ratio for infants in a given state might be much higher than the ratio for older children, but our explanation for ratios in excess of 100 percent suggests that if the ratio exceeds 100 percent among infants, it should do so for older children as well.

In sum, this comparison suggests that the high group test provides an effective means of identifying SSNs that are not in the Numident at the time. One must be careful, though, that the high group codes used in the test are contemporaneous with the data. Valid SSNs assigned after the reference period for the high group codes used in the test may be deemed invalid.[17]

2.   Reasons for Passing or Failing Validation

As we have explained, SSNs on enrollment records can pass the Census Bureau's validation tests without complete agreement between the enrollment record and the Numident with respect to gender and date of birth. Similarly, when Mathematica unduplicated MAX records in earlier research, perfect agreement on demographic characteristics was not required to define a pair of records with the same Medicaid identification number (MSIS ID), SSN, or HIC as the same individual. It is of interest, therefore, how often the records with SSNs that passed the Census Bureau's validation tests had varying levels of agreement with the Numident on gender and date of birth. Furthermore, if we imposed a more rigorous standard than the Census Bureau employs, how much would the validation rate decline, and how would this vary by state?

It turns out that the records that passed the Census Bureau validation tests almost always agreed completely on gender and date of birth. Over all age groups 97.55 percent of the MSIS records with validated SSNs had the same gender and date of birth as the Numident (Table II.13). For all but the elderly, full agreement on these characteristics was even higher, ranging from 97.74 percent for non-elderly adults to 98.49 percent for infants. For elderly

---

[17] In addition, the high group test will not be applicable to SSNs issued after June 2011.

adults, full agreement was nearly three percentage points lower, at 94.67 percent. Of the seven additional patterns of agreement that were allowed, the next most common for all but infants was agreement on gender and year of birth and one (but not both) of the month and day of birth (pattern 4). For elderly adults this combination occurred in 2.88 percent of the validated records but was much lower at younger ages. The second most common combination for infants—and third most common for all but the elderly—was full agreement on the date of birth but not gender (pattern 6). This pattern occurred in 0.76 percent of the validated infant records and 0.67 percent overall. Among elderly adults, the second most common pattern was agreement on gender and both the month and day of birth but a discrepancy of one year on the year of birth (pattern 2). Not surprisingly, this pattern increased in frequency with rising age.

When Mathematica unduplicated MAX records, the last four patterns in Table II.13 would not have been counted as agreement when records were matched on SSN rather than the MSIS identifier (Verghese and Czajka 2012). Pattern 5, with agreement on gender and year of birth but the month and day of birth inverted, would have been a reasonable addition to the MAX unduplication, but this pattern accounted for only 0.07 percent of the records that passed validation. Pattern 8, which included agreement on gender and either the month or day of birth with a discrepancy of one year in the year of birth, could have been included in the MAX unduplication as well, but this pattern was even less common, occurring in only 0.05 percent of the validated records. The two remaining patterns included disagreement on gender, which was not allowed in the MAX unduplication on CMS's recommendation. We discussed pattern 6 above. Pattern 7 involved agreement on the year of birth, with the month and day inverted and disagreement on gender. This pattern had too few occurrences to register in the table (fewer than 0.005 percent).

Table II.13. Patterns of Agreement on Gender and Date of Birth: Medicaid Enrollees with Validated SSNs by Age, Fourth Quarter, FY 2009

Pattern of Agreement	Infants (< 1)	All Other Children (1 to 18)	Non-elderly Adults (19 to 64)	Elderly Adults (65+)	Total
1. Gender and year, month, and day of birth	98.49	97.96	97.74	94.67	97.55
2. Gender and month and day of birth, but year differs by 1	0.04	0.11	0.27	1.08	0.27
3. Gender and month and day of birth, but year differs by more than 1	0.03	0.03	0.10	0.50	0.11
4. Gender and year of birth, and either month or day of birth but not both	0.66	1.03	1.24	2.88	1.29
5. Gender and year of birth, but month and day are inverted	0.00	0.04	0.10	0.10	0.07
6. Year, month and day of birth but not gender	0.76	0.80	0.53	0.55	0.67
7. Year of birth, but month and day are inverted; not gender	0.00	0.00	0.00	0.00	0.00
8. Gender and either month or day of birth (but not both), and year differs by 1	0.01	0.03	0.03	0.22	0.05

Source: State MSIS files, fourth quarter, FY 2009.

There was very little variation among the states in the proportion of validated SSNs with complete agreement on gender and date of birth. Full agreement was lowest among elderly adults in every state, ranging from 88.5 to 96.7 percent (Table II.14). For each of the other three age groups, the proportion with full agreement was distributed over a narrower range—from 96.1 to 99.5 percent for infants; from 95.4 to 99.1 percent for older children; and from 95.0 to 98.9 percent for non-elderly adults.

When SSNs that were in the Numident failed validation, there was little or no agreement on gender and date of birth. We did not estimate all of the possible ways that these fields could disagree, but selected patterns are reported in Table II.15. In almost one percent of the cases (but

Table II.14. Percentage of Validated SSNs with Agreement on Gender and Full Date of Birth, by Age and State: Fourth Quarter, FY 2009

State	Infants (< 1)	All Other Children (1 to 18)	Non-elderly Adults (19 to 64)	Elderly Adults (65+)	Total
United States	98.5	98.0	97.7	94.7	97.5
Alabama	97.4	97.6	98.1	92.8	97.1
Alaska	97.8	97.8	97.5	93.4	97.4
Arizona	99.3	98.4	97.9	95.7	98.0
Arkansas	98.5	97.6	98.2	93.2	97.4
California	98.7	98.2	97.6	96.7	97.8
Colorado	98.1	98.0	98.1	95.1	97.7
Connecticut	98.5	98.2	97.7	93.4	97.4
Delaware	98.4	98.2	97.6	92.6	97.5
District of Columbia	97.1	96.5	95.0	88.5	95.1
Florida	98.6	98.4	98.1	95.6	97.9
Georgia	98.1	97.9	98.0	92.3	97.3
Hawaii	98.9	98.7	98.3	96.3	98.3
Idaho	98.7	98.2	98.5	96.3	98.2
Illinois	97.9	96.2	96.8	92.6	96.2
Indiana	99.2	99.0	98.5	95.1	98.6
Iowa	98.1	98.2	98.5	95.6	98.1
Kansas	99.1	98.9	98.6	95.6	98.5
Kentucky	96.3	95.4	98.0	92.9	96.1
Louisiana	98.1	98.0	98.2	93.1	97.6
Maine	99.0	98.8	98.8	96.1	98.3
Maryland	97.4	97.4	97.0	93.1	96.9
Massachusetts	98.5	97.8	97.3	94.5	97.2
Michigan	99.1	98.3	98.5	95.5	98.2
Minnesota	99.0	98.6	98.5	96.6	98.4
Mississippi	98.2	98.2	97.5	91.2	97.1
Missouri	99.4	99.1	98.5	94.7	98.5
Montana	99.3	98.8	98.6	95.5	98.5
Nebraska	99.0	98.1	98.5	95.6	98.0
Nevada	98.1	97.3	97.8	96.5	97.4
New Hampshire	99.0	98.5	98.4	95.2	98.1
New Jersey	96.1	96.3	96.9	93.6	96.1
New Mexico	97.5	98.0	97.7	94.2	97.6
New York	98.4	97.4	96.5	94.3	96.6
North Carolina	99.2	98.6	98.3	92.9	98.0
North Dakota	99.5	99.0	98.8	96.1	98.6
Ohio	98.6	98.5	98.3	94.0	98.1
Oklahoma	98.8	97.6	97.7	94.8	97.4
Oregon	97.9	97.6	98.2	96.2	97.7
Pennsylvania	98.7	98.3	97.9	93.7	97.7
Rhode Island	98.9	98.6	97.8	93.5	97.7
South Carolina	97.4	97.2	97.2	90.5	96.6
South Dakota	99.0	98.6	98.5	95.6	98.3
Tennessee	99.1	98.7	98.2	93.9	98.1
Texas	98.5	98.0	98.2	95.0	97.8
Utah	99.1	98.6	98.4	96.2	98.5
Vermont	98.2	98.1	98.4	95.6	98.0
Virginia	98.0	97.4	97.4	92.7	96.9
Washington	98.1	97.7	98.3	96.5	97.8
West Virginia	99.0	98.6	97.6	93.9	97.8
Wisconsin	99.3	98.9	98.9	96.5	98.7
Wyoming	97.4	97.5	98.3	95.5	97.5

Source: State MSIS files, fourth quarter, FY 2009.

Table II.15. Patterns of Agreement on Gender and Date of Birth: Medicaid Enrollees with Numident SSNs That Failed Validation, by Age, Fourth Quarter, FY 2009

Pattern of Agreement	Infants (< 1)	All Other Children (1 to 18)	Non-elderly Adults (19 to 64)	Elderly Adults (65+)	Total
1. Numident date of birth is incomplete	0.18	0.19	0.81	3.34	0.89
2. Gender and year of birth, but both month and day disagree (and are not inverted)	11.95	11.81	17.98	32.90	17.00
3. Gender, but the year of birth differs by 1, and both month and day disagree	3.32	3.28	3.52	8.50	4.25
4. All else	84.54	84.72	77.69	55.26	77.86

Source:　　State MSIS files, fourth quarter, FY 2009.

3.34 percent of elderly adults) the Numident date of birth was incomplete, which the Census Bureau treated as missing. Situations where gender and the year of birth agreed but month and day did not (and were not inverted) represented 17.00 percent of the failed validations (but nearly a third among elderly adults, as this pattern increased with age). A similar pattern but with the year of birth differing by one accounted for another 4.25 percent of the validation failures (but 8.50 percent among the elderly). More than three-quarters of the failures fell into an "all else" category. We did not examine the reasons for failure by state, given that such failures occurred for only half a percent of the SSNs that were submitted for validation and that the patterns of disagreement were not particularly informative.

C.　Changes in Reporting and Validity, 2007 to 2009

If SSN reporting has improved in recent years, then there exists the prospect of continuing improvement. If, on the other hand, SSN reporting has not improved, then we would have good reason to anticipate that improvements in the near future are unlikely. To assess recent changes in the reporting of SSNs in MSIS records, we compared rates of reporting of SSNs in the valid range for MSIS enrollees with full benefits between two time periods: the first quarter of FY 2007 (October through December 2006) and the fourth quarter of FY 2009 (July through

September 2009). We focused on beneficiaries with full benefits because recipients of restricted benefits are often not required to have (or report) SSNs. Improvements in the low reporting rates for SSNs among beneficiaries with restricted benefits would depend on changes in state policies rather than better collection practices.

Over this period of nearly three years, the percentage of MSIS records with SSNs in the valid range among enrollees with full benefits increased by 0.6 percentage points, from 96.7 percent to 97.3 percent (Table II.16). Among the states, which are the source of any improvements in the quality of SSN reporting, the change between the two years was uneven. A total of 27 states exhibited an increase in the percentage of SSNs in the valid range while 3 showed no change and 21 showed a reduction. The largest increases were gains; Arkansas and Massachusetts experienced improvements just over 8 percentage points, which brought the reporting rates in these two states from around 90 percent to 98 percent or higher, which was above the national average in 2009. After these two states, the next largest improvement was recorded by Delaware, where the reporting rate for SSNs in the valid range increased by 3.1 percentage points, from 92.3 percent to 95.4 percent.

Five other states had SSNs in the valid range for fewer than 95 percent of their full benefit enrollees in the first quarter of 2007. Of these five states, three showed slightly improved reporting in the fourth quarter of 2009 (Georgia, Nevada, and New Jersey), but the other two showed small reductions (Montana and New York).

Based on these patterns, we would expect smaller improvements over the next three years than we observed between the first quarter of FY 2007 and the fourth quarter of FY 2009. The two states with the lowest reporting rates showed large improvements, so they can exhibit only small gains in the future. The next five states showed mixed results. While they continue to

Table II.16. Percentage with an SSN in the Valid Range: Medicaid Enrollees with Full Benefits by State, First Quarter, FY 2007, and Fourth Quarter, FY 2009

State	First Quarter FY 2007	Fourth Quarter FY 2009	Change
United States	96.7	97.3	0.6
Alabama	97.8	96.4	-1.4
Alaska	96.0	97.2	1.2
Arizona	98.6	98.8	0.2
Arkansas	90.3	98.6	8.3
California	95.1	95.6	0.5
Colorado	95.4	96.8	1.4
Connecticut	96.2	96.4	0.2
Delaware	92.3	95.4	3.1
District of Columbia	98.0	96.1	-1.9
Florida	98.5	99.1	0.6
Georgia	94.8	95.6	0.8
Hawaii	98.0	98.9	0.9
Idaho	97.8	96.7	-1.1
Illinois	98.5	98.9	0.4
Indiana	99.3	97.0	-2.3
Iowa	99.4	98.8	-0.6
Kansas	99.4	99.1	-0.3
Kentucky	99.0	99.0	0.0
Louisiana	97.8	95.2	-2.6
Maine	99.6	99.2	-0.4
Maryland	97.4	96.3	-1.1
Massachusetts	89.9	98.0	8.1
Michigan	95.3	98.0	2.7
Minnesota	97.9	99.1	1.2
Mississippi	95.8	97.3	1.5
Missouri	98.6	98.5	-0.1
Montana	94.6	93.1	-1.5
Nebraska	98.6	98.5	-0.1
Nevada	92.3	93.7	1.4
New Hampshire	99.3	99.1	-0.2
New Jersey	94.2	94.9	0.7
New Mexico	98.7	98.3	-0.4
New York	94.9	94.6	-0.3
North Carolina	98.7	98.4	-0.3
North Dakota	99.8	99.8	0.0
Ohio	98.9	97.9	-1.0
Oklahoma	97.8	97.9	0.1
Oregon	97.1	97.5	0.4
Pennsylvania	99.5	99.3	-0.2
Rhode Island	100.0	99.1	-0.9
South Carolina	95.9	97.5	1.6
South Dakota	98.2	98.3	0.1
Tennessee	99.6	99.4	-0.2
Texas	95.8	96.9	1.1
Utah	98.6	99.2	0.6
Vermont	99.8	99.8	0.0
Virginia	95.1	96.3	1.2
Washington	97.0	97.5	0.5
West Virginia	100.0	99.4	-0.6
Wisconsin	97.4	99.1	1.7
Wyoming	96.1	97.4	1.3

Source: State MSIS files, first quarter, FY 2007, and fourth quarter, FY 2009.

have room for significant improvement, their recent experience suggests that large improvements in the near term are not likely. The remaining states showed mixed results as well, and given that these states had markedly high reporting rates for the most part in the fourth quarter of FY 2009, we would not expect more than a modest improvement over the next few years.

Among SSNs in the valid range, the fraction passing Census Bureau validation was already close to 100 percent in the first quarter of 2007, leaving little room for growth, and it grew only marginally through the fourth quarter of 2009. That it grew at all should perhaps be considered surprising. Nationally, the fraction passing validation rose from 99.2 percent to 99.4 percent, an increase of 0.2 percentage points (Table II.17). States with relatively low validation rates in the first quarter of 2007 had the most room to grow, and they showed the most improvement. The most dramatic increase occurred in Arkansas, where the validation rate improved from 98.7 percent to 99.8 percent. Virginia showed the next largest increase, rising from 98.6 percent to 99.4 percent. Both Colorado and Rhode Island registered increases of 0.6 percentage points. Only Massachusetts recorded a reduction, dropping very slightly from 99.5 percent to 99.4 percent. We would expect little change nationally over the next few years, but in seven states the validation rates were still below 99 percent in the fourth quarter of 2009. These states have the most need and the most potential for improvement, but we note that only two of these states (Illinois and New Jersey) showed increases in excess of the national rate of 0.2 percentage points between the first quarter of FY 2007 and the fourth quarter of FY 2009.

Table II.17. Percentage of SSNs Passing Census Bureau Validation: Medicaid Enrollees with an SSN in the Valid Range by State, First Quarter, FY 2007, and Fourth Quarter, FY 2009

State	First Quarter FY 2007	Fourth Quarter FY 2009	Change
United States	99.2	99.4	0.2
Alabama	99.2	99.3	0.1
Alaska	98.9	99.3	0.4
Arizona	99.0	99.3	0.3
Arkansas	98.7	99.8	1.1
California	99.4	99.5	0.1
Colorado	98.5	99.1	0.6
Connecticut	98.6	98.8	0.2
Delaware	98.4	98.5	0.1
District of Columbia	98.3	98.5	0.2
Florida	98.8	99.2	0.4
Georgia	98.5	98.7	0.2
Hawaii	99.7	99.8	0.1
Idaho	99.9	99.9	0.0
Illinois	98.4	98.9	0.5
Indiana	99.8	99.8	0.0
Iowa	99.3	99.4	0.1
Kansas	99.8	99.8	0.0
Kentucky	99.6	99.7	0.1
Louisiana	99.1	99.2	0.1
Maine	99.3	99.4	0.1
Maryland	99.1	99.1	0.0
Massachusetts	99.5	99.4	-0.1
Michigan	99.6	99.7	0.1
Minnesota	99.5	99.7	0.2
Mississippi	99.1	99.5	0.4
Missouri	99.7	99.9	0.2
Montana	99.8	99.8	0.0
Nebraska	99.8	99.9	0.1
Nevada	99.5	99.5	0.0
New Hampshire	99.3	99.4	0.1
New Jersey	97.6	97.9	0.3
New Mexico	99.3	99.4	0.1
New York	99.1	99.4	0.3
North Carolina	99.5	99.7	0.2
North Dakota	99.8	99.9	0.1
Ohio	99.2	99.3	0.1
Oklahoma	99.4	99.5	0.1
Oregon	98.8	99.1	0.3
Pennsylvania	99.3	99.4	0.1
Rhode Island	99.0	99.6	0.6
South Carolina	98.9	99.2	0.3
South Dakota	99.1	99.2	0.1
Tennessee	99.7	99.7	0.0
Texas	99.1	99.2	0.1
Utah	99.6	99.8	0.2
Vermont	99.6	99.7	0.1
Virginia	98.6	99.4	0.8
Washington	99.4	99.8	0.4
West Virginia	99.7	99.8	0.1
Wisconsin	99.3	99.8	0.5
Wyoming	98.8	98.8	0.0

Source: State MSIS files, first quarter, FY 2007, and fourth quarter, FY 2009.

III. DISCUSSION AND CONCLUSION

Valid SSNs were present on 91.7 percent of the MSIS records submitted by the states for the fourth quarter of 2009—the latest quarter for which such data were available from all of the states. SSNs were missing on 7.7 percent of the MSIS records and incorrect on another 0.6 percent of the records, according to tests performed with SSA data maintained and enhanced by the Census Bureau. When an SSN was reported, it was almost always correct: 99.3 percent of the reported SSNs were judged to be valid.

Validation at the Census Bureau was based on agreement on gender and date of birth between the MSIS record and the Numident file, which contains data from applications for SSNs. While agreement on these characteristics did not have to be perfect for an SSN to be considered valid, it nearly always was. Across all age groups, 97.55 percent of the valid SSNs were on MSIS records with the same gender and date of birth as the corresponding Numident record. Another 1.29 percent agreed on all but the month or day of birth, and an additional 0.27 percent disagreed only on the year of birth, which differed by one. Together these patterns accounted for 99.11 percent of the valid SSNs.

The MSIS records do not contain names or addresses, so the validation was not quite as rigorous as it could have been if such information had been present. If an MSIS enrollee "borrowed" an SSN from another person of the same gender and approximately the same age and reported the SSN owner's date of birth, the validation test would not detect an error. If such SSN sharing were so sophisticated, however, it is possible that names might be borrowed as well, in which case a more rigorous validation might yield the same result. Rather than focusing on such speculative possibilities, we think it is more important to consider the implications of our findings at their face value.

The exceedingly high validation rate for MSIS SSNs is significant to CMS because the validation results for individual Medicaid enrollees cannot be removed from the Census Bureau. The Numident data are covered by Title 13 of the federal code, which means that individually identifiable data from the Numident cannot be shared outside of the Census Bureau.[18] Even without this information, the high validation rate allows CMS to treat the reported SSNs as essentially correct in all cases. Further, some of the incorrect SSNs—although not many—can be identified with the high group test. The implications are especially good for applications involving record linkage between MSIS and other databases.

Lastly, our research at the Census Bureau was undertaken because SSA was reluctant to engage in a process of validating the high volume of MSIS SSNs. SSA performs such functions for federal and state agencies on a smaller scale and has developed computer programs to perform validation based on names as well as gender and dates of birth. In addition to working with a high volume of records, performing validation for MSIS SSNs would involve modifying programs in order to perform a less rigorous procedure. Using a dedicated server at the Census Bureau, we were able to validate 50 million SSNs in about half an hour. We do not know how this translates into run time at SSA and what other limits the SSA staff may face in scheduling time on SSA's mainframe computers, but it would seem that Mathematica's experience at the Census Bureau would be useful to share with SSA if CMS remains interested in validating SSNs at the individual level. However, with a validation rate of 99.4 percent among SSNs that fall into the valid range, a minimum validation rate of 97.9 percent among the states, and only six other states below 99 percent, micro-level validation would appear to offer limited marginal value. It is difficult to imagine an application where serious error could result from simply assuming that

---

[18] To perform tabulations within a secure computing environment at the Census Bureau, Mathematica staff had to obtain special sworn status.

all reported SSNs are correct. For applications involving record linkage, use of the high group test or other appropriate range checks prior to linkage would eliminate most of the illegitimate SSNs (those that were never assigned), and matching on gender and date of birth in addition to SSN would prevent nearly all matches to MSIS records with incorrect SSNs.

This page has been left blank for double-sided copying.

REFERENCES

Czajka, John, and Jenna Libersky. "Expenditures and Service Use Associated with MAX Records without SSNs." Memorandum to William Clark, David Baugh, and Cara Petroski. Centers for Medicare & Medicaid Services, May 10, 2011.

Czajka, John L., Audra Wenzlow, and Julie Sykes. "Development of the Medicaid Analytic Extract Enrollee Master (MAXEM) File, 2005 and 2006." Final Report. Washington, DC: Mathematica Policy Research, September 14, 2010.

Verghese, Shinu, and John L. Czajka. "Linking Medicaid Administrative Records over Time and Space: Methods and Findings, 2005 to 2007." Presented at the 2012 Federal Committee on Statistical Methodology Research Conference, Washington, DC, January 2012.

This page has been left blank for double-sided copying.

MATHEMATICA
Policy Research

Improving public well-being by conducting high-quality, objective research and surveys

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC

Mathematica® is a registered trademark of Mathematica Policy Research