

HCFE Data Brief

HCFE DB# 2008-01

Technical Documentation: Medicaid Enrollment, Utilization and Outcomes for VA Patients, V1.0

- Understanding MSIS data
- Creating Analytic Variables Using Medicaid, Medicare and VA Data

John A. Gardner
Ann M. Hendricks
Lynn Wolfsfeld

June 2008



Health Care Financing & Economics

Boston VA Healthcare System, 150 South Huntington Avenue, Mail Stop 152H, Boston, MA 02130
• Phone: (857) 364-6058 • Fax: (857) 364-4511 • Web: www.hcfe.research.va.gov

TABLE OF CONTENTS

1. Introduction	3
2. Medicaid Data	4
3. Creation of Study Variables	13
4. References	18
5. Appendix A	19
6. Appendix B	26

This work was funded by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Services (ECI 03-199). This report presents the findings and conclusions of the authors and does not necessarily represent VA or HSR&D.

1.0 INTRODUCTION

In 2004, the Veterans Health Administration (VA) Health Services Research & Development (HSR&D) funded “Medicaid Enrollment, Utilization and Outcomes for VA Patients (ECI 03-199) to pursue three major objectives:

1. Relate the use of VA, Medicaid and Medicare services nationally for all veterans dually enrolled in VA and Medicaid (VA+Medicaid) to the patients’ needs.
2. Identify Medicaid program factors leading to greater VA+Medicaid enrollment.
3. Compare risk-adjusted outcomes for VA+Medicaid enrollees and comparable VA-only patients.

To meet these objectives, the research team combined Medicaid and Medicare data for calendar years 1999 and 2000 with the matching months of VA utilization from fiscal years 1999, 2000 and 2001 (October 2000 through December 2000 are in FY2001). While many VA researchers have experience working with merged VA and Medicare data, this project was the first in VA to use the national Medicaid datafiles. This report of some of the methods required to work with these different data sources is intended as background for other papers from this project and as a guide for other researchers who may use Medicaid data for their projects.

The report has two major sections. The first describes the Medicaid data in some detail. The second presents information on creating analytic variables using the Medicaid, Medicare and VA files. We illustrate the data issues through two examples: assigning VA patients to a particular geographic location and counting patients’ visits with a clinical provider.

2.0 MEDICAID DATA

Since 1999, the Centers for Medicare and Medicaid Services (CMS) have maintained Medicaid data in two different formats - MSIS Medicaid Statistical Information System (MSIS) and Medicaid Analytic eXtract (MAX). This report deals primarily with the first of those data formats, the MSIS version of national Medicaid eligibility, utilization and payment data, because the research team used MSIS for this project under a data use agreement with CMS. The research team chose to obtain and use MSIS data because of the relative timeliness of the MSIS data compared to the MAX data. At the time we requested data, MAX data for the necessary years was not yet available. The reason we include MAX files in this discussion is that individual VA Researchers are able to, at no cost, obtain MAX files by submitting a data request to the Veterans Information Research Center (VIREC). Under a Memorandum of Understanding with CMS, VIREC obtains both Medicare and Medicaid data directly from CMS and redistributes it to VA researchers. At the time of this writing VIREC has still not completely finalized the data request process for obtaining MAX files. It is also unclear to the authors whether VA researchers can directly request MSIS files from CMS. The next two subsections describe the structure of the MSIS and MAX files, with an emphasis on the differences between them. Because MAX files are derived from MSIS files, some of the MSIS issues discussed in this document pertain to the MAX files as well.

2.1.1 MSIS files

The MSIS datafiles are submitted by the states to CMS on a quarterly basis. They include records for claims that were “paid, processed, voided or otherwise adjudicated”¹ in some manner by the state Medicaid agency during that quarter, even if the services were actually

performed in another (necessarily earlier) quarter. For example, the files for the third quarter (July through September) of 1999 may have records for services provided in that quarter, in earlier quarters of 1999, and even from earlier calendar years. Because MSIS data files are organized by quarter, data can be obtained for whatever combination of quarters is desired to make up either a fiscal year or a calendar year.

The research team obtained CMS MSIS data from several sources. Under data reuse agreements with JEN Associates, Brandeis University and Columbia University, we obtained MSIS eligibility and claims data for CY1999 through CY2000 1st quarter as well as pharmacy files for CY1999. Under a data use agreement with CMS, we received CY2000 2nd-4th quarter and 1st quarter CY2001 eligibility and claims data, and CY2000 1st quarter – CY 2001 1st quarter pharmacy data. In total we obtained 9 quarters of data. The 5 additional quarters of data beyond CY1999, were necessary to capture total utilization occurring in CY1999.

Datafiles were provided on 3490E magnetic cartridges. For each state, for each quarter, there were separate tapes for eligibility and for each claim type. For some especially large states there were multiple tapes per claim type per quarter. In all, CMS provided approximately 1200 tapes with data for the entire United States population of Medicaid enrollees for the requested time period. Initially, several states, some of them very large and therefore important to the analysis had failed to submit all the required quarterly files (NY, PA, WI, NV). Eventually, we acquired these files as well. In order to subset the population of veterans we contracted with JEN Associates of Cambridge, Massachusetts to process the data. We provided JEN Associates with a VA finder file of approximately 8.4 million VA patients identified as having utilized VA from 1995 -2001. We were interested not only in obtaining data for veterans who were utilizing both VA and Medicaid during 1999-2000, but also in identifying veterans who might sequentially have utilized VA and then Medicaid services (e.g., veterans who utilized VA in 1997 and 1998, did not utilize VA in 1999, but did utilize Medicaid in 1999). JEN Associates conducted the match using social security number (SSN), validating the match by checking gender and date of birth, an approach that its staff has used extensively to match Medicaid and Medicare files and that VA researchers have also used to match Medicare and VA data.² The match produced 425,665 patients some of whom were non-veterans. After excluding non-veterans, we identified 377,896 as veterans. In looking at 407,000 patients who were identified as eligible in 1999 approximately 20% were not VA patients in CY1999 or CY2000.

2.1.2 MAX files

The Medicaid Analytic eXtract (MAX) files – formerly known as State Medicaid Research Files (SMRFs) – are a set of claim-level data files derived from MSIS data and include eligibility, service utilization and payments. The MAX files abstract the records for services performed during a given calendar year from seven quarters of submitted claims. The choice of seven quarters of MSIS data to construct one year of MAX data was made by CMS to balance completeness of data and timeliness of MAX file availability. “Per CMS, when this issue (of completeness) was last reviewed, they found a range of greater than 99% for the pharmacy (RX) file to about 95% for the Inpatient (IP) claims. The rank order from most complete to least would be RX, Other Therapeutic (OT), Long Term Care (LT) and IP. This is only in terms of fee-for-service claims. The capitated care encounter claims are not felt to be very complete and weren't part of this assessment.”³

MAX files enhancements include: adding eligibility data to each claim, placing retroactive eligibility in the proper order, combining all interim claims so that only final action

events are recorded, organizing the data by dates of service, including additional eligibility such as dual Medicare and Medicaid status, QMB and SLMB status, Medicare HIC number, (Temporary Assistance to Needy Families (TANF) and State Children’s Health Insurance Program (SCHIP) eligibility among others.⁴ Data limitations for both MSIS and MAX files include: missing payments, third party payments may not be accounted for, drug payment amounts listed are prior to rebates and no provider characteristics are available.⁴

MAX Files are available on CD, DVD or USB hard drive depending on the size of the file. More information on the history, structure and availability of MAX files can be found at <http://www.resdac.umn.edu/Medicaid/medicaidFAQ.asp>. The ResDAC site also includes links to MSIS and MAX data dictionaries as well as pdf documents listing data anomalies by state. Additional general information on the MAX files is available at the CMS website http://www.cms.hhs.gov/MedicaidDataSourcesGenInfo/07_MAXGeneralInformation.asp

2.2 Description of MSIS Data

MSIS datafiles contain two basic types of records: eligibility records and claim records. Claim records include four types of claims - inpatient, long term care, other care and prescription drugs. This subsection provides an overview of the MSIS records including the data included in each type of record and the limitations of the data.

2.2.1 Eligibility Records

Eligibility records are quarterly – three months of information per record [January through March, April through June, etc.]. The monthly indicators can show the reason for that person’s Medicaid eligibility, whether the enrollee has private health insurance paid for by a third party, whether the enrollee is enrolled in a managed care plan (and if so, for what broadly-defined kinds of services).

Only the Eligibility records have a field for SSN. As a result, these records are absolutely necessary for matching a finder file. Even so, MSIS does not guarantee that the field contains an SSN unless the state requires it.

According to the MSIS Tape Specifications released in March 2000,⁵ CMS mandates that eligibility records contain:

- one record for each person who was eligible for Medicaid for at least one day during the reporting quarter covered by the file or who, at State option, is being included as a non-Medicaid CHIP record;
- one record for each individual for whom retroactive eligibility was established during the reporting quarter and for each prior reporting quarter covered by the retroactive eligibility;
- corrections to ELIGIBLE File records submitted in prior quarters. Note: All correction records must be submitted as complete records. Do not submit records that contain valid values only in the corrected fields. Correction records will completely replace the eligible record previously provided.

An alphabetic list of the variables in the Eligibility Records with a brief description of each is provided in Table 1.

Table 1: Eligibility Records

Field	Information contained
-------	-----------------------

Basis of eligibility	Aged, blind/disabled, unemployed, foster child, other
Chip code	Whether in a Medicaid CHIP program
County code	FIPS county code of residence
Date of birth	Date
Date of death	Date
Days of eligibility	Days eligible for Medicaid in month
Dual eligible flag	Medicare eligibility (QMB, SLMB, other)
Eligibility group	Combination of basis of eligibility and maintenance assistance status
Federal FY, quarter	Federal fiscal year and quarter to which record applies
Health insurance status	Whether had private health insurance paid for by state or by a third party
HIC number	Medicare Health Insurance Claim Identification Number
Income code	Family's state-defined income bracket (state bracket specifications not given)
Maintenance assistance status	Medically needy, poverty related, Section 1931, Section 1115 demonstration, or other basis
MSIS case number	State-assigned MSIS case number (not a personal identifier, and a given person's case number may change over time)
MSIS ID number	Unique personal identifier. For SSN states, may have a temporary ID, if one assigned. (SSN given in separate field.) For non-SSN states, contains a unique personal identifier assigned by the state.
Plan-ID (up to 4)	Specifies managed care plans under which eligible person is covered
Plan type	Medical or comprehensive medical, dental, behavioral, prenatal/delivery, long-term care, PACE, primary care case management, other
Race/ethnicity	White, Black or African American, American Indian or Alaskan Native, Asian, Hispanic or Latino, Hawaiian Native, multiple race, other
Restricted benefits flag	Whether full benefits or limited because of alien status, pregnancy status, Medicare status, special Medicaid program
Sex	Male, female
SSN	Social security number
TANF-cash flag	Temporary Assistance for Needy Families
Type of record	Record gives current quarter info, retroactive quarter, or correction or update of earlier quarter
Zip code	Zip code of residence

There are limitations and potential problems in working with MSIS eligibility data. Some of those that this project team had to resolve include:

1. Multiple MSIS IDs exist for a given SSN. For CY1999 and 2000, we found roughly 546,000 MSIS IDs for 431,000 SSNs. Initially, all MSIS IDs corresponding to a

particular SSN were kept. This allowed researchers to link all claims data to an SSN and keep track of the linkage. At later points in the study process, where possible, differences were resolved using age and gender as described below.

2. A given SSN may appear in multiple states (and therefore may be associated with multiple MSIS IDs. For our study years, 0.7% of the 431,000 SSNs appeared in the files for two states in any half year. In order to assign a patient to a state, numbers of months of eligibility in each of the multiple states were compared and the state for which patients had the most months of eligibility recorded was assigned as the patient's state. For example, a patient with 4 months of eligibility in Maine and 2 months in Massachusetts was assigned to Maine (row 4). Table 2 shows examples of instances when patients had eligible months in multiple states. Freq1 shows the number of months of eligibility in State1 and Freq2 shows the number of months of eligibility in State2. The Assigned state is the state for which the patient had the most eligible months. Each row corresponds to one patient.

Table 2: Assignment of state using greatest number of eligible months

Assigned state* (Larger of Freq1 or Freq 2)	State1	State2	Freq1 (# of months eligibility in State 1)	Freq2 (# of months eligibility in State2)
ME	NH	ME	6	11
AZ	AZ	CO	15	5
NH	NH	OR	4	2
ME	ME	MA	4	2
MA	MA	NH	16	6
ME	ME	FL	5	3
VT	VT	NH	15	8
CA	CA	VA	15	9
ME	ME	LA	20	4
NM	ME	NM	9	11
NH	NH	WA	15	9
GA**	GA	FL	10	10
CA	CA	AZ	17	6
ME	TX	ME	3	9
ME	ME	PA	17	2
NC	ME	NC	4	7

*Dominant state based on state with highest number of records in two years.

**In cases where the frequencies were tied, the state was assigned based on other criteria such as the VA ZIPcode, location at the start of the study period or another reason determined by the study objectives.

3. Eligibility records may be updated from one quarter to the next.
 - 3-5% of records in CY1999-2000 data are updates
 - 1% of eligibility records in any year update information for a different CY
 - About 75,000 out of 2.1 million records are updates within the CY

This complexity of the data requires extra care on the part of the data handler. If a quarter contains both a standard entry and an update, the researcher should use the update. If there are more than one update, use the last one.

4. Conflicts across records in age or gender for a given SSN
 - Among MSIS IDs, approximately 1% had conflicting values, even after incorporating updated records
 - Among SSNs, about 0.5% (3,100 out of 431,000) had conflicts

For both of these conflicts, we used data on gender and age from the Medicare denominator file where available; otherwise we used VA data on gender and age to resolve the conflicts.

Other issues

The field labeled ‘dual-eligibility flag’ purports to flag enrollees who are in both Medicaid and Medicare. This field is probably not very reliable, especially for early years of MSIS data.⁵ According to JEN Association, studies sponsored by CMS in specific states using detailed Medicaid data maintained by the state itself have shown that there is considerable discrepancy between actual Medicare enrollment, as known to the state, and Medicare enrollment in CMS files for MSIS.⁵ VA researchers can use Medicare data from VIREC to establish whether the VA patient had joint Medicaid and Medicare eligibility or enrollment in any given study period.

For all our VA-Medicaid enrollees, 25.3% were enrolled in Medicaid for all 24 months in calendar years 1999 and 2000. No other tally of months from 1 through 23 exceeded 5% [6] of the study population. Table 3 lists the number of months of eligibility, the number of patients eligible for that number of months and the percent of patients who were eligible for specific numbers of months.

See Appendix A for a discussion of additional issues.

*Table 3. Number of Months of Eligibility,
All VA Patients, FY1999-FY2000*

Month	Frequency	Percent
Missing	1,377	0.32
1	19,786	4.65
2	19,419	4.56
3	25,082	5.89
4	16,449	3.86
5	16,266	3.82
6	20,378	4.79
7	14,664	3.44
8	13,053	3.07

9	16,422	3.86
10	11,972	2.81
11	11,621	2.73
12	17,483	4.11
13	10,492	2.46
14	9,220	2.17
15	11,477	2.7
16	8,500	2
17	8,428	1.98
18	20,991	4.93
19	6,745	1.58
20	7,096	1.67
21	17,225	4.05
22	6,809	1.6
23	7,102	1.67
24	107,608	25.28

*Patients eligible in 2 states are counted using the sum of all eligible months.

Completeness of data is an ongoing issue. States are required to submit MSIS data on a quarterly basis, but at times are late in filing. As filing has shifted from a voluntary to mandatory program, submissions are more often completed in a timely manner.

2.2.2 Claims Records

Claims records have an MSIS ID number, which is also in the Eligibility records (see Table 1). In many states, this MSIS ID number is the SSN or contains the SSN. The MSIS ID is the only way to link eligibility or claims records to each other. There is no state ID provided in the MSIS data. When we had our data processed by JEN Associates, JEN modified The MSIS ID to include a state indicator based on the quarterly state file the ID was contained in. We recommend this step for others working with this data who wish to track state trends. This also makes it possible to differentiate IDs for a single patient who gets benefits in two states, if both states use the SSN as the MSIS ID number.

MSIS includes four types of claims files:⁶

1. Inpatient Claims (CLAIMIP), includes acute inpatient care
2. Long Term Care Claims (CLAIMLT), includes nursing facilities, intermediate care facilities, psychiatric hospitals and the psychiatric wards of acute-care hospitals
3. Other Therapy Claims (CLAIMOT), outpatient services and services provided to inpatients by providers located outside the inpatient facility (e.g., private duty nurses, physicians), also payments for capitated services or those received under a Title XIX waiver
4. Prescription Drug Claims (CLAIMRX).

CMS mandates that each claim file tape submitted to CMS:⁶

- contain one record for every claim of the appropriate type paid, or encounter processed, during the reporting quarter and
- conform to one of the four standard claims file record formats and data element lists, although many data elements are common to all four claims files.

Note that inclusion of a record in a particular calendar quarter of MSIS data depends on the date that the state pays or processes the claim, not on the date of service. This configuration contrasts with the MAX data, which organizes claims by date of service. See Table 4 below for tabulations of the service dates of claims in the MSIS data for calendar years 1999 and 2000. The columns showing the proportions of claims and amounts paid within the year of service demonstrate that the average lags for paying acute inpatient care are the longest; for prescription claims, the shortest.

Table 4: Timing of Claims Data in MSIS data

Type of Claim	Calendar Year	Number of Claims		Amounts Paid (Absolute values)	
		Count	% within calendar year (apportioned)	Dollars (Millions)	% within calendar year
Inpatient	1999	194,507	65.8	469.9	63.2
	2000	232,191	72.7	635.9	64.7
LTC	1999	1,207,120	74.0	1,581.2	72.5
	2000	1,957,104	70.9	2,255.9	73.0
Other	1999	13,340,567	76.7	851.8	77.4
	2000	19,292,755	79.6	1,120.3	78.4
RX	1999*	8,000,945	93.8	329.2	95.3
	2000	11,326,228	94.9	416.8	95.6

* Do not have any Hawaii data, New York data for quarters 3 and 4, or fourth quarter data for Pennsylvania, Wisconsin or Nevada.

Claims files are supposed to include:

- one record for each line item that is separately adjudicated;
- all fully adjudicated current quarter claims that have completed the State's processing cycle, for which the State has determined that it has liability to reimburse the provider;
- all adjustments to prior quarter claims adjudicated in the reporting quarter;
- adjudicated claims that passed all the States' eligibility and coverage edits, but resulted in a zero liability because of payments by responsible third parties;
- claim records representing capitated payments or fees paid to capitated plans;
- encounter claims (TYPE-OF-CLAIM=3), to the extent that they are routinely received by the State;
- Medicare/Medicaid Crossover claims, which are identified by the presence of valid values in the Medicare-DEDUCTIBLE-PAYMENT and Medicare-COINSURANCE-PAYMENT fields.

MSIS submission should not include any claim that does not relate to covered Medicaid services (e.g., claims for services to non-Medicaid CHIP individuals), or that has been returned

to the provider because of insufficient information. Table 5 is an alphabetic list of the variables in the Claims Records

Table 5: Claims Records

Field	Information contained
Adjustment indicator	Whether original claim or adjustment to prior claim record
Admission date (IP, LTC)	Date of admission
Amount charged	Charge on claim
Beginning date of service	Date of service covered in claim, or first of multiple dates, or start of period covered
Date of payment adjudication	Date on which state determined claim payment status
Date prescribed (RX)	Date
Days supply (RX)	Days of medication dispensed
Diagnosis code (principal) (IP)	Condition determined to be chiefly responsible for admission, even if another diagnosis is involved that is more severe
Diagnosis code (up to 9)	ICD-9-CM. Up to 2 on other, 5 on LTC
DRG (IP)	Diagnosis related group
DRG group indicator	Grouping algorithm used to set DRG
Ending date of service	Analogous to beginning DOS
ICF – MR days (LTC)	Every day of ICF-MR paid by state, even if part of day is paid for by others
Leave days (LTC)	Days covered by Medicaid during which the patient was not resident in LT facility
Medicaid amount paid	Amount paid by Medicaid
Medicaid covered IP days (IP,LTC)	Days of covered IP, LT or neonatal care
Medicare coinsurance payment	Amount paid by Medicaid toward patient's Medicare coinsurance
Medicare deductible payment	Amount paid by Medicaid toward patient's Medicare deductible payment
MSIS ID number	See eligibility file description
NDC (RX)	National drug code
New refill indicator (RX)	Whether new prescription or refill
Nursing facility days (LTC)	Days of nursing care paid for (in whole or in part) by Medicaid, including leave days
Other third-party payment	Amount paid by source other than Medicaid, Medicare, patient
Patient liability (LTC)	Total amount to be paid by patient
Patient status	Discharged, transferred, died, left against medical advice, admitted as inpatient, still a patient
Place of service (OT)	Place of service (from HCFA 1500)
Plan ID number	Number of health plan under which service

	in claim was provided
Prescribing physician ID (RX)	Identifying number
Prescription fill date (RX)	Date medication dispensed
Procedure code (principal) (IP)	CPT, ICD-9, HCPCS code (or state-specific coding structure [no detail])
Procedure codes 2 thru 6	Other procedure codes, same possible coding structures listed for principal
Proc code flags (1 thru 6)	Identifies coding structure that applies to the specific procedure listed
Proc code mod (1 thru 6)	Modifiers for procedure code
Proc date (principal)	Date principal procedure performed
Program type	Special Medicaid program if applicable (EPSDT, family planning, rural clinic, FQHC, Indian health services, HCBC)
Provider ID billing number	ID for specific provider that billed for service covered in claim
Servicing provider ID number	ID for provider giving service (other than billing provider)
Quantity of service (OT, RX)	Number of units of service provided
Service code (OT)	CPT, ICD-9, HCPCS code (or state-specific coding structure [no detail])
Service code flag	Identifies coding structure that applies to the specific procedure listed
Service code mod	Modifier for service code
Specialty code (OT)	(No standard code, only state-specific codes [no detail given])
Type of claim	Current fee-for-service, capitated payment, dummy record simulating a bill for service under some capitation plans, service-tracking record for lump-sum payments, supplemental payment
Type of service	IP, ICF, physician services, lab or x-ray, capitated payments, etc.
UB-92 Revenue code	For UB-92 line item
UB – Revenue charge (up to 23) (IP)	Amount charged
UB – Revenue units (up to 23) (IP)	Units

For inpatient services, the record allows entry of up to six procedure codes. There is also a revenue center code for each service. These fields are often blank, however. In the LTC and Other claims records, HCPC or CPT codes are used to identify the service that was provided to the patient. The coding scheme varies from state to state and may even vary from claim to claim within a state. Some states may use a state-specific coding scheme that is not used elsewhere. In these cases, the coding structure may not be easily obtained. Thus, there is frequently ambiguity as to who provided service and where. Each state has its own cross-walk for provider IDs. Further, the “place of service” entries are often not very precise or informative. There is no certainty that states interpret the codes the same way as to type of facility or provider.

The prescription records have space for the date of the prescription, the date filled, and the date paid, but not all of these fields are always populated. Similarly, the field for the days of supply for the medication is not always filled. To identify the specific drug, Medicaid includes only the National Drug Code (NDC), not the name. This characteristic makes it somewhat difficult to compare the prescriptions under Medicaid with the medications dispensed through VA, which records the name of the medication but has less reliable capture of the NDC.

3.0 CREATION OF STUDY VARIABLES

This section presents information about the creation of variables for the Medicaid project. These descriptions not only provide methods background for other papers for this project, they may help other researchers create their own variables from multiple data sources. The two examples we focus on are a geographic location variable and a visit count variable.

3.1 Geographic Location

Assigning individuals to states is necessary to perform certain analyses. First, assigning patients to a particular state is important because it is necessary to match patients to specific state Medicaid programs in order to control for the characteristics of the Medicaid programs (e.g., generosity of benefits) at the level of the state. This match is needed not only for Medicaid enrollees, but also for any comparison group not in Medicaid. Secondly, assigning patients to a specific ZIPcode is important to estimate the distance a VA patient has to travel to get to a VA or an alternate source of care.

The research team anticipated that some inconsistencies would arise for the same individual when comparing residence as recorded separately by VA, Medicaid and Medicare. For example, an individual might be enrolled in Medicaid in one state but have a ZIPcode in another state for their VA utilization. To focus on state's Medicaid rules and patient choices, the following hierarchical approach for assigning residency status to an individual is suggested:

- 1) If for a specific time period, a valid ZIPcode is recorded in the Medicaid data, assign that ZIPcode as the state of residence.
- 2) If for a specific time period no valid ZIPcode exists in the Medicaid data, and for that same time period a valid ZIPcode appears in the Medicare data, assign the Medicare ZIPcode as the determinant of the state of residence.
- (3) If no valid ZIPcode appears in either the Medicaid or Medicare files for the time period, look to adjacent time periods in the Medicaid files to see if there is a valid ZIPcode and take that code if it exists.
- (4) If no ZIPcode exists in adjacent Medicaid files, and there is a valid VA ZIPcode for the time period, apply the VA ZIPcode. If there are multiple VA ZIPcodes, apply the one that appears most often.
- 5) If there is no ZIPcode in VA inpatient data for the time period, look to see if there is a valid ZIPcode in VA outpatient data.

Using these 5 steps project staff were able to assign 99.8% of Medicaid patients and 91-92% of patients who were not on Medicaid to states. For the remaining, unassigned patients, further steps included looking at adjacent time periods and non adjacent time periods of the relevant datasets. See Appendix B for the complete algorithm.

3.2 Clinician Visits

For our Medicaid analyses, we used clinical visits to measure the degree to which dual enrollees relied on the VA for care. Each source of information about health care utilization (VA, Medicare, Medicaid) has a different way of identifying visits with a clinical provider. This subsection describes the methods we used to count clinical encounters in VA and in the non-VA claims datafiles.

VA Codes

In the VA, researchers may use the Clinic Stop code to identify when a patient has met with a clinical provider. Table 6 lists the codes for 49 clinic stops where there are patient provider interactions. This set of clinic appointment types was chosen because they are 1) appointment types with high volumes, 2) appointment types that cover 93% of office-based patient-provider interactions in the VA (versus other services, such as labs or telephone consultations) and 3) appointment types that represent all major sub-specialties of medicine (e.g. mental health, orthopedics).⁷

Table 6: VA Clinic Stop Codes Used to Define Clinical Provider Visits

Stop Code	Description	Stop Code	Description
123	Nutrition/diet	322	Women's clinic
149	Radiation therapy	323	Primary care
201	Physical Medicine and Rehabilitation Services (PM &RS)	350	Geriatric primary care
202	Recreation therapy	401	General surgery
203	Audiology	403	Enterology
204	Speech pathology	404	Gynecology
205	Physical therapy	406	Neurosurgery
206	Occupational therapy	407	Ophthalmology
210	SCI	408	Optometry
214	Kinesiotherapy	409	Orthopedics
301	General internal medicine	410	Plastic surgery
302	Allergy Immunology	411	Podiatry
303	Cardiology	413	Thoracic surgery
304	Dermatology	414	Urology
305	Endocrine/metabolism	415	Vascular surgery
306	Diabetes	420	Pain clinic
307	Gastroenterology	502	Mental health – individual
308	Hematology	509	Psychiatry - individual
310	Infectious disease	510	Psychology – individual
312	Pulmonary/chest	513	Substance abuse – individual
313	Renal/nephrology	523	Opioid substitution
314	Rheumatism/arthritis	531	Mental health primary care - individual
315	Neurology	540	PTSD team – PCT
316	Oncology	547	Intensive substance abuse treatment
317	Anti-coagulation clinic		

Medicare Codes

In Medicare claims, outpatient care with a clinical provider may be recorded in either the physician supplier or the institutional outpatient files. For physician supplier claims, we considered each line item record; for institutional outpatient claims, each revenue center record. The codes to identify clinical visits in the Medicare claims are listed in Table 7.

Table 7: Medicare Codes Used to Identify Clinical Visits

Variable Name	Code	Description
Revenue Center	360-379	Operating room, anesthesia
	450-469	ER/pulmonary functions
	480-529	Cardiac, ambulatory surgical center, OP services
	900-919	Psychologic services
	961-963	Professional fees, psychological
	973-975	Professional fees, radiation
Place of Service	11	Office
	12	Home
	22	OP hospital
	23	ER
	24	Ambulatory surgical center
	50	Federal health center
	53	Community mental health center
	62	Comprehensive OP rehabilitation center
	65	ESRD
	71	State or local public health center
	72	Rural health clinic
Type of service	1	Medical care
	2	Surgery
	3	Consultation
	6	Therapeutic radiation
	7	Anesthesia
	T	Psychological therapy

Medicaid Codes

Clinical providers are recorded in the “Other Therapy” claims records for the Medicaid programs. To be counted as a clinical encounter in our analyses, the record had to have both a specified place of service (as indicated in Table 8) and the “Type of service” as a physician. Unlike Medicare, there are no revenue center codes in Medicaid. Additionally, while Medicare codes clinical encounters with non-physicians, Medicaid codes only clinical encounters with physicians (or with dental providers which we chose not to include in our analysis).

Table 8: Medicaid Codes Used to Identify Clinical Visits

Variable Name	Code	Description
Place of Service	11	Office
	12	Home

	22	OP hospital
	23	ER
	24	Ambulatory surgical center
	50	Federal health center
	53	Community mental health center
	62	Comprehensive OP rehabilitation center
	65	ESRD
	71	State or local public health center
	72	Rural health clinic
Type of Service	08	Physician

Counting Encounters

For both Medicare and Medicaid claims, there is ambiguity about the number of visits that a specific claim covers. A given claim may cover a period of many days. Typically, a claim lists beginning and ending dates covered. While these may be the same date, indicating a single unit of each type of visit, for a substantial number of claims, these dates are not the same, indicating that the services could have been provided over a period of time (e.g., a week or month). There is also a field that is variously labeled “quantity of service” in Medicaid claims, “Units” in the Medicare outpatient revenue center detail records or “Service count” in the physician supplier line item detail records. This field often has values that are greater than one. For example, about 10% of the Medicare Physician/Supplier line-item records in our data had a ‘Service count’ value greater than 1.

Considering each record to represent only 1 visit would thus substantially undercount the number of provider visits in the data. It is clear, however, that the units or service count field often contains values greater than 1 even when the beginning and ending service dates are the same. As a result, treating units or service count as the number of visits recorded in the datafile could overcount non-VA visits by a substantial amount.

After reviewing Medicare and Medicaid claims, we settled on an algorithm for counting visits that compromises between the extremes of under- and overcounting. If the beginning and ending dates for a record were the same, we counted the record as showing 1 visit. If the beginning and ending dates differed, then we took the number of visits to be the lower of two values: the service count (or units) field or a number derived from the differences in the two dates. If the beginning and ending service dates were within 60 days of each other, we assumed that visits would be weekly, that is the rounded integer value equal to $\lceil \frac{\text{difference in dates}}{7} \rceil$. If the beginning and ending service dates differed by more than 60 days, we assumed that visits would be monthly: rounded integer value equal to $\lceil \frac{\text{difference in dates}}{30} \rceil$. To summarize, the count of visits from a given Medicare or Medicaid claim would be 1 (if beginning date = ending date), or the lesser of the units reported or the weekly or monthly visits calculated from the difference between the beginning and ending service dates.

Because there was also the possibility that both Medicare and Medicaid would be reporting the same visits (if Medicare was the primary payer), we excluded Medicaid records that indicated either a co-payment or deductible that Medicaid would have to pay for a claim otherwise covered by Medicare. We also excluded Medicaid claims that showed a negative amount paid by Medicaid, because these claims indicated a refund to Medicaid for service that was provided (and recorded) in another claim with the same dates of service.

Descriptive Means

Defining the number of visits using the approach described above and excluding those patients for whom there were more than 90 visits in a 6 month period resulted in overall counts that ranged from zero to 177 per person in 1999 and zero to 179 in 2000. The ranges for VA visits were zero to 177 and zero to 179 for the study years. For Medicaid, the counts ranged from zero to 169 and zero to 166. For Medicare they ranged from zero to 172 and 178.

Despite the large variation in the number of clinical visits in the three programs, the mean numbers of visits varied little across a number of variables that VA researchers are used to including as factors to explain utilization differences, including VA priority status, age, having nursing home stays, psychiatric co-morbidities or by enrollment in a Medicare HMO. Overall, patients averaged 11 visits a year, roughly one a month. The VA accounted for about half of the average and Medicare visits averaged more than twice those under the Medicaid program.

REFERENCES

1. www.resdac.umn.edu/Medicaid/FAQs-draft-08172006.doc
2. VA Information Resource Center, Research Findings from the VA Medicare Data Merge Initiative: Veteran's Enrollment, Access and Use of Medicare and VA Health Services, Report to the Under Secretary for Health, Department of Veterans Affairs, September 2003.
3. www.resdac.umn.edu ResDAC Request 27670, May 29, 2008 2:27:18 PM
4. Baugh, Dave. Medicaid Analytic eXtract: An Introduction. Centers for Medicare & Medicaid Services (CMS) Office of Research, Development and Information (ORDI) Updated July 13, 2006 <http://www.cms.hhs.gov/MedicaidDataSourcesGenInfo/Downloads/maxintro.pdf>
5. Oral communication with JEN Associates, 2002.
6. Medicaid Statistical Information System (MSIS) TAPE SPECIFICATIONS AND DATA DICTIONARY. Release, 2, Version 2.1. Updated February 2000. Released March 2000. Effective Fiscal Year 1999. HCFA, Center for Medicaid and State Operations and Office of Information Services.
7. Baar, B. 2005a. "New Patient Monitor: Data Definitions". Veteran Health Administration Support Services Center.

APPENDIX A

Screening Medicaid eligibility data to identify unique patients and match with VA patients.

MSIS eligibility datafiles are not very clean. There are duplicate records, updates of records, multiple states for a given patient, multiple MSIS IDs for a given SSN, logical inconsistencies in the recorded basis for Medicaid eligibility across multiple records for a given MSIS ID, and records that clearly apply to dependents, not to the veteran. This memo describes steps taken to try to identify VA patients themselves.

I Get the latest eligibility information for a given MSIS ID.

MSIS eligibility records specify a calendar quarter: Jan-Mar, Apr-Jun, etc. Each record has separate entries for each month of the quarter. The entries describe the basis for eligibility, the maintenance assistance status, the types of managed care plans, and the identity of specific managed care plans (in state specific codes for which the documentation does NOT have a key). The codes for “Basis-of-Eligibility”, “Maintenance-Assistance-Status”, and type of managed care plan (PLAN-TYPE-1, PLAN-TYPE-2, PLAN-TYPE-3 and PLAN-TYPE-4) are delineated in the Eligible File section of the MSIS Tape Specs and Data Dictionary descriptions of those variables.

The datafiles have 3 types of records: a standard entry, a retroactive entry, and an update or correction of a previous entry. For example, a retroactive entry would be a record that was submitted in Q3 of FY1999 but whose information applied to a beneficiary’s eligibility in Q1 of FY1997. The entry is retroactive if there has never been a record submitted to MSIS for that beneficiary for Q1 of FY1997. If the beneficiary did have a record in MSIS describing eligibility status for Q1 of FY1997, then the new record (submitted in Q3 of FY1999) would be an update or correction of a previous entry. (This report always refer to CY, not FY, for these data, but the actual submissions to MSIS identify quarters in reference to federal fiscal years. This project identified the timing of eligibility only in terms of a FY and quarter, not directly as a CY quarter. The datafiles tire provided as a collection of quarters that comprise a calendar year.)

The CY99 eligibility file has 1,271,178 records, 16,253 of which were retroactive entries; 24,226 were update/correction records.

In trying to specify the latest eligibility information for a given MSIS ID:

- 1) Collect all entries for a given MSIS ID. Sort by fiscal year, quarter, and type of record.
- 2) For a given quarter, take the record with the highest record type (corresponding to the order listed above). Take an update/correction record, if one exists, otherwise a retroactive record. Take a standard entry only if there is no update/correction record for that quarter. [The datafile should not have both a retroactive record and a standard entry record.]
- 3) If a given quarter has more than one update/correction record, take the last one in the dataset. The assumption here is that the records are accumulated in chronological (by quarter) sequence of submission to MSIS. If the actual submission quarter for a record was known, then one would not have to rely on this assumption to choose the latest

record; one could be confident that one had the latest record. Because each tape that MSIS provides contains one quarter of data for a state, it should be possible to retain that information in the construction of an eligibility file.

This screening process reduced the eligibility file from 1,271,178 records to 1,253,382 records.

II Identify all of the MSIS IDs that apply to a given beneficiary.

The project identified VA patients by their SSN. States know a beneficiary’s SSN, and include it in the eligibility data, but there are two classes of states: 1) the SSN becomes the Medicaid ID, 2) there is a Medicaid ID distinct from the SSN. States with 1) are called ‘SSN states’; states with 2) are called ‘Other’ or ‘Non-SSN’ states. The MSIS ID contains the Medicaid ID plus an indicator of which state the ID is for. As a result, it is possible to differentiate IDs for a single patient who gets benefits in two states, even if both are SSN states.

Although MSIS eligibility datafiles have a field that lists SSN, MSIS claims data records are identified by the MSIS ID, not the SSN itself. With SSN states, it is possible to match a claim directly to a SSN (using part of the MSIS ID); with non-SSN states, that cannot be done directly. One has to determine the MSIS ID that covers the SSN, then screen claims against the MSIS ID.

A single SSN may be associated with more than one MSIS ID, meaning that some beneficiaries have MSIS IDs in more than one state. Presumably, there are conditions under which this kind of duplication can be legitimate, but multiple-matching also occurs sometimes within a single state. The conditions under which this one-to-many relationship is legitimate are not clear and the MSIS documentation does not say anything about this kind of multiple ID.

With the MSIS ID as the unit of analysis, there are not too many anomalies. After following the rule described above to limit the data to one record per quarter per person, the months of eligibility are as shown in Tables 1-2.

TABLE 1: Frequency of Eligible Months

eligmnts	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	2338	0.56	2338	0.56
1	22076	5.29	24414	5.85
2	21136	5.06	45550	10.91
3	50797	12.16	96347	23.07
4	16225	3.89	112572	26.96
5	15704	3.76	128276	30.72
6	17480	4.19	145756	34.91
7	14556	3.49	160312	38.39
8	13611	3.26	173923	41.65
9	32350	7.75	206273	49.40
10	12404	2.97	218677	52.37
11	12578	3.01	231255	55.38
12	186317	44.62	417572	100.00

TABLE 2: Frequency of Months Not Eligible

months	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	279094	66.84	279094	66.84
1	58468	14.00	337562	80.84
2	64332	15.41	401894	96.25
3	9508	2.28	411402	98.52
4	4433	1.06	415835	99.58
5	869	0.21	416704	99.79
6	532	0.13	417236	99.92
7	114	0.03	417350	99.95
8	59	0.01	417409	99.96
9	86	0.02	417495	99.98
10	5	0.00	417500	99.98
11	16	0.00	417516	99.99
12	56	0.01	417572	100.00

A small number of MSIS IDs have no eligibility during CY1999. Their entries must arise from retroactive or update records that apply to quarters outside CY1999.

Otherwise, the selection process described above does, in fact, pare the data down to no more than 4 eligible (quarterly) records and a maximum of 12 eligible months. For 451 MSIS IDs, the eligibility codes for one or more months were ‘9’ or ‘unknown’. Counts of eligible months and not eligible months do not appear to reconcile because months that fall in quarters for which there is no eligibility record are not counted as ‘not eligible’ months. The counts shown are of months for which an eligibility status is recorded in an eligibility record.

For any given MSIS ID, the information is highly consistent. Only 2 MSIS IDs had two or more SSNs listed in eligibility files for that single MSIS ID. There were no instances of conflicting gender or multiple dates of birth for a given MSIS ID.

When SSN is the unit of analysis, things get messy. There are 417,572 distinct MSIS IDs. There are 9,143 SSNs that have more than a single MSIS ID associated with them. Altogether, these SSNs are linked to 18,912 MSIS IDs, just slightly over 2 per SSN. For one SSN, there are 16 (quarterly) eligibility records in the data. Another 17 SSNs have 5 or more eligibility records.

One clue to the source of these multiple MSIS IDs comes from looking at the state that is linked to the MSIS ID. Florida has the most, with 2,725 of these IDs. This number exceeds even California (1,776) and New York (1,095). It would appear that alien status or changes in citizenship status may be associated with having multiple IDs. The ‘snowbird’ phenomenon may also contribute.

One possible explanation for multiple IDs associated with the same SSN is that the MSIS eligibility is for a dependent, not the VA patient, and that the VA patient’s SSN is associated with the dependent’s MSIS record. The dependent would most likely be a child, though it could sometimes be a spouse.

The basis of eligibility variable has 3 codes that indicate children:

- '4' – child (not '6' or '8')
- '6' – child of unemployed adult
- '8' – foster child

The data show 8,286 MSIS IDs for which at least one month of eligibility has one of those codes. Of these, 7,477 show one or another of these codes in all of their eligible months. That is, there is no month with an affirmative eligibility code that specifies a non-child basis for eligibility. That leaves 809 IDs for which there is at least one month recorded as aged, blind, unemployed or adult(not unemployed) as the basis for eligibility. Approximately 300 of those IDs show a child eligibility code in fewer than half of the months for which they show affirmative eligibility status.

Apparently, all of those MSIS IDs for which all eligible months consistently indicate 'child' status should be excluded from analysis, as they should not be veterans. However, the DOB information in the MSIS data is not always consistent with the BOE info. For about 200 of these 'children', the DOB implies an age over 22; for over 100, the implied age exceeds 37. Nine are indicated as 65 or older. Clearly, there is some inconsistency in the demographic and eligibility information for some of these beneficiaries.

For 8,187 of these beneficiaries, either all months of eligibility show a 'child' code or the age implied by the date of birth falls between 1 and 24. For another 41 beneficiaries, at least half the months of affirmative eligibility show 'child' status. For the remaining 58 IDs, the case for the beneficiary being a child is not convincing. Compare age/gender information for these beneficiaries with the age/gender information from VA for the associated SSN. Make any decisions to include or exclude these beneficiaries based on those comparisons.

To pin down possible spousal status for these IDs would involve steps to identify the age and gender of the VA patient associated with the SSN, then compare those to the MSIS information about the beneficiary. It appears that 'child' status may account for most cases of multiple MSIS IDs.

III Inconsistencies between VA information and MSIS information.

Both VA data and MSIS data identify the date of birth (DOB) and gender of the patient. Researchers should check the extent to which those identifications are consistent with each other. If gender and/or DOB differ, there is reason to doubt that the people being identified in the VA and MSIS data are the same person.

As noted above, for a given MSIS ID, the gender and DOB information are consistent across eligibility records. For VA data, there may be greater variability across utilization records in the specification of DOB or gender. However, in accumulating VA information, VA researchers will likely try to pick the most frequent identification. In a few cases, there could be correct information in MSIS data and wrong information in VA data, but VA researchers are unlikely to affirm that condition with administrative data. Where there is a difference between VA

information and MSIS information, decide whether the difference is substantial enough to conclude that the MSIS claim data are not referring to the same person as the VA data.

With gender, the issue of agreement between data sources is straightforward. If the reported gender differs, the data sources disagree. However, that disagreement may not always dictate disregarding the MSIS data for the person. If the DOB data agreed exactly, then one might be tempted to believe that the data referred to the same person, despite the disagreement on reported gender. Same dates of birth for any two people chosen at random will seldom agree. Combine that with the apparent equality of reported SSN in two data sources, and it becomes easier to suspect that one or the other source has simply recorded the wrong gender in its automated data. One of the decisions to make is whether to override differences in reported gender to treat the two patients as being the same, but such disagreement in gender is not too frequent (see below).

With DOB, there is more range than with gender for treating two different entries as applying to the same person. There are three elements to a date: year, month and day. One of those elements might be miscopied, misread, or mis-keypunched.

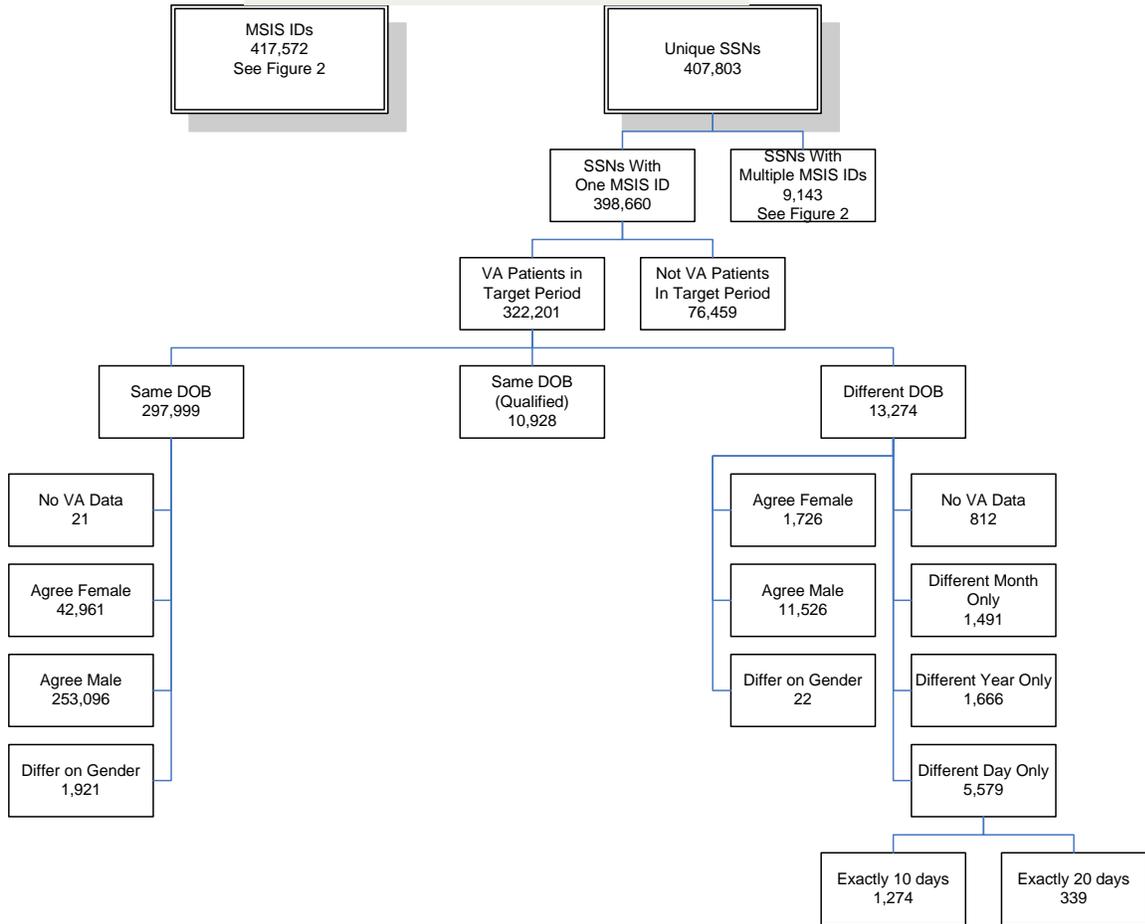
The MSIS eligibility information for CY1999 contained 417,572 MSIS IDs, which corresponded to 407,803 distinct SSNs. Because there are some SSNs that have more than one MSIS ID, this report separates the comparison process into two parts. Part one compares VA and MSIS data only for those SSNs that have a single MSIS ID (Figure 1). Part two examines SSNs for which there is more than one MSIS ID (Figure 2).

To what extent might these disagreements on gender or DOB be attributable to dependents being recorded in MSIS instead of the VA patient himself/herself? It appears that dependents account for, at most, a small portion of the disagreements in data.

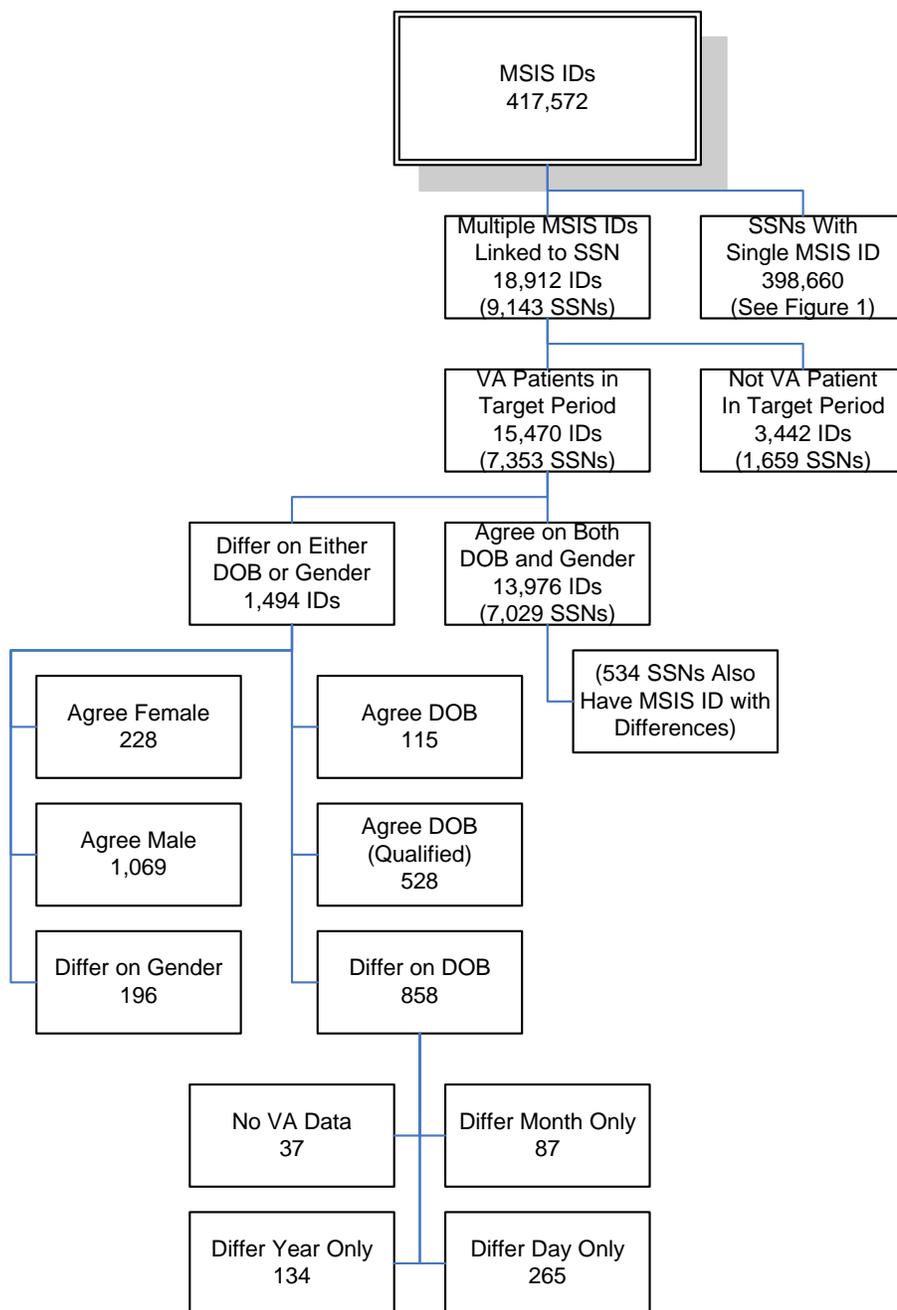
Consider first the SSNs for which there is only a single MSIS ID. For those, of the 13,274 records that have some disagreement on DOB, only 266 correspond to MSIS IDs that are identified above as involving children. Of the 1,989 records that disagree on gender, 121 of them are child records.

Of the 13,976 SSNs for which there are multiple MSIS IDs and agreement on both DOB and gender, 362 are child records. For the 1,494 for which there is disagreement on DOB or gender, 128 are child records. Of the 3,442 for SSNs that were not VA patients in the target period, 82 of them were child records.

**Figure 1 – VA/MSIS Data
Compare Gender and DOB
SSNs With One MSIS ID**



**Figure 2 – VA/MSIS Data
Compare Gender and DOB
Multiple IDs Per SSN**



APPENDIX B – ALGORITHM FOR ZIPCODE ASSIGNMENT

The algorithm operates as follows:

- 1) Within the half-year, if there is a valid MCD ZIPcode, take it [MCD 1]. (Indicate if more than one code was available in the MCD data for that half-year. [MCD 2] Recall that there are 2 quarterly observations in each half-year, and for a few SSNs there are 2 MCD IDs)
- 2) Otherwise, if there is a valid MCR ZIPcode in the calendar year denominator data, take that. (If the MCR zip equals a VA zip for the same half-year, indicate so. Also, if the state implied by the MCR ZIPcode is not equal to the state listing from the MCR patient address, indicate so. Equal VA zip and MCR zip state = MCR listed state: MCR E. Equal VA zip but MCR states differ: MCR N. Not equal VA, MCR states same: MCREV. Not equal VA, MCR state differ: MCRNV)
- 3) Otherwise, if there is a valid MCD ZIPcode in an adjacent half-year (only 99b possible for this 99a selection), take it. [MCDA1] (As in 1), indicate if >1 zip available [MCDA2])
- 4) Otherwise, if there is a valid zip shown in VA inpatient data for the half-year, take it. Take the most frequently occurring VA zip if more than one available. [VOK] If there is a tie among two or more zips for most frequent, take first but indicate tie in the condition code [VTIE]. If the most frequent zip is shown only once, indicate that [V_1_].
- 5) Otherwise, follow 4) for VA outpatient data for the half-year. (same codes)
- 6) Otherwise, follow 4) for VA inpatient, then outpatient data for adjacent half-years. [VAOK, VATIE, or VA_1_]
- 7) Otherwise, check non-adjacent half-years for a MCD zip. Process as in 3) MCDB1 or MCDB2]
- 8) Otherwise, check other year's denominator file for MCR zip. Process as in 2) [MCRAE, MCRAN, MRAEV, MRANV]
- 9) Otherwise, check non-adjacent half-years for VA IP, then OP ZIPcode. Process as in 4) [VBOK, VBTIE, VB_1_]
- 10) Otherwise, set condition code = 'NONE', state = ' '.