

# Data Investigation

## Vital Status File Date of Birth (DOB) and Sex Quality Review

### Background

The VHA Vital Status File (VSF) combines information from multiple VA and Non-VA data sources to provide the most complete information available on Veterans' vital status and date of death. The Vital Status Master file has one record for each SSN-date of birth-sex combination found in VHA and VBA source files [1]. Incomplete date of birth (DOB) and/or sex as well as conflicting DOB and/or sex from different data sources will result in multiple records for a single SSN on the VSF. Because researchers use the SSN, date of birth, and sex to match Veteran data in study cohorts to the VSF to obtain the Veteran's vital status and date of death, we assessed the consistency and accuracy of the DOB and sex data in the VSF. Specifically, we evaluated the:

1. Extent and source of incomplete or missing DOB and sex information;
2. Extent and causes of multiple records, i.e., missing DOB and/or sex versus conflicting DOB and/or sex;
3. Rates of concordance between VA DOB and sex and Social Security Administration (SSA) or Centers for Medicare and Medicaid Services (CMS) data, two sources of demographic data generally considered reliable.

More in-depth information about the VSF is available on the VIREC website [1]. For convenience, we use the term "demographics" in this document to refer to DOB and/or sex information generally.

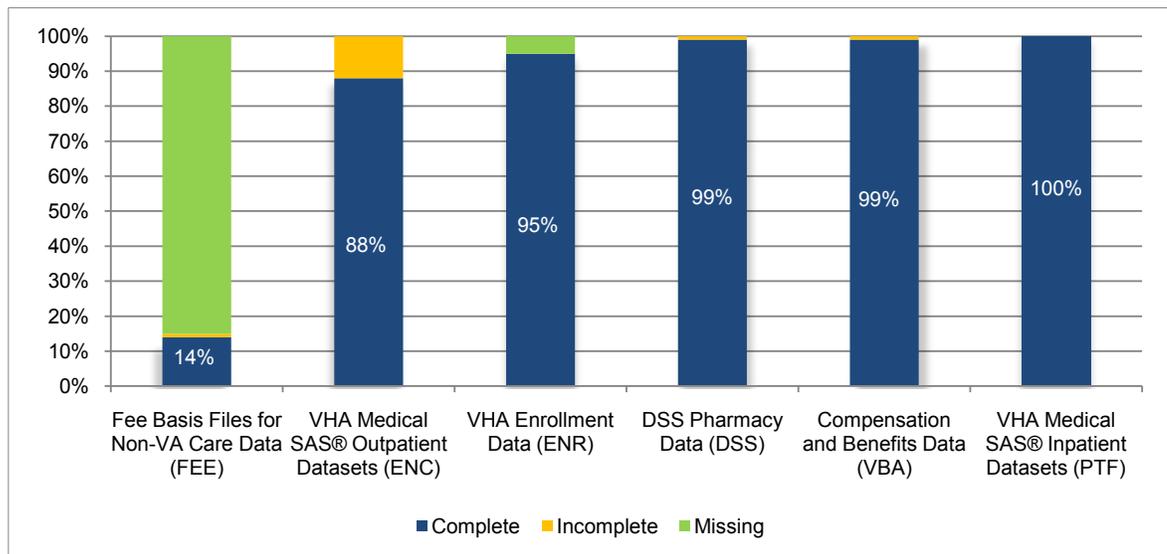
The VSF Master created on December 12, 2010 was used in this analysis. This file contained 21,982,516 records of which 19,869,512 were Veteran records. There were 14,009,035 unique SSNs on the Veteran records. Only Veteran records were used in this analysis.

## 1. Incomplete or Missing DOB and Sex Information

For each SSN in the VSF Master, DOB and sex information may originate from six sources within the VA: the VHA Medical SAS® Outpatient Datasets (ENC); VHA Medical SAS® Inpatient Datasets (PTF); Fee Basis files for non-VA care data (FEE); DSS pharmacy data (DSS); VHA enrollment data (ENR); and compensation and pension benefits data (VBA) extracted from the Veterans Benefits Administration VETSNET database. A single DOB-sex combination on the VSF may originate from multiple sources. The DOB and sex may be entirely missing, incomplete, or complete in the source data. For example, the sex may be recorded as missing or unknown and DOB may contain the year only. We examined the completeness of the DOB (i.e., 3 parts, DDMMYYYY) and sex (i.e., present/missing) information from each source.

**Missing data.** Dates of birth provided by most VHA sources had high rates of completeness. The exception was FEE; less than 11% of SSNs from the FEE source had 3 parts DOB and 85% had no DOB. Other rates of completeness ranged from 88% (ENC source) to 100% (PTF source). Figure 1 shows the completeness of DOB information from the six VA sources.

**Figure 1. Completeness of DOB Information by VA Source**



Sex was missing in the FEE source data 85% of the time. Rates of missing sex values from other sources ranged from 0 (PTF source) to 5% (ENR source).

**Defaulted values.** Every source file DOB on the VSF has a corresponding variable that indicates whether the DOB is complete (i.e., has all 3 parts populated). Note that a complete date may have 'defaulted' values when the true DOB is unknown. Default values are often "1" in the day position and "January" in the month position. For example, if a DOB contains "December 1, 1945", the day may have been set to "1" in the source database because the day of birth was unknown or the day of birth may actually be the first. In this example, the complete-DOB indicator would indicate a complete DOB in either case. Based on a comparison of the percentage of the VSF dates of birth with the day set to "1" and/or the month set to "January" versus the percentage that would be expected if every day of the year had an equal number of births, we estimated that approximately 5% of the dates of birth had defaulted values in the month and/or day. This percentage varied by source of the date of birth and was greatest for the VBA source. A previous analysis of dates of birth on the VETSNET file suggests that approximately 12% of the VBA dates of birth might have an unknown month and/or day of birth set to the first day of the month or first day of the year [2].

## 2. Multiple Records

Of the 14,009,035 unique Veteran SSNs on the Master file, 8,902,772 (63%) had a single record and 4,415,436 (32%) had two records (i.e., two different DOB-sex combinations found in the source data). The remaining 690,827 (5%) SSNs had three or more DOB-sex combinations. One SSN had 22 different records.

Multiple records due to missing or incomplete DOB and sex. As noted above, an additional record will be generated for an SSN when DOB and/or sex values are different, even if the only difference is owing to a missing or incomplete value.

Among the 5,106,263 SSNs with more than one DOB-sex combination, 3,196,615 (63%) had multiple records only because the DOB and sex were both missing from one or more sources. When records missing both DOB and sex were removed (including 10,112 SSNs this record as the only record), 12,089,275 (86%) of the 14,009,035 unique Veteran SSNs on the file had a single (non-missing) DOB-sex combination in the data (8,902,772 + 3,196,615 – 10,112).

An analysis of the percentage of Veterans with multiple records by age was performed for a subset of Veterans alive on January 1, 2010 (i.e., no recorded date of death or the date of death was after December 31, 2009) using the Veteran's age on July 1, 2011. The percentage of Veterans with multiple records was similar in Veterans 65 years of age and older (34%) versus those under 65 (39%). When records with DOB and sex both missing were excluded the percentages of SSNs with multiple records dropped to 9.8% for Veterans 65 and older and 8.4% for Veterans under 65.

Multiple records due to conflicting information. Among the remaining 1,919,760 SSNs with multiple records, 1,199,045 (63%) had conflicting DOB and/or sex and 710,603

(37%) had no conflicting information but the DOB or sex from one or more sources was incomplete. Thus, 8.6% of all SSNs for Veterans on the Master file had conflicting dates of birth and/or sex values recorded in the VHA and VBA source data (1,199,045 out of 14,009,035 SSNs). Of the SSNs with conflicting demographics, 2.2% had conflicting DOB and sex, 14.7% had conflicting sex only, and 83.1% had conflicting DOB only.

Table 1 details the conflicting demographic information by age and DOB.

**Table 1. Conflicting Demographic Information on the VSF Master File**

Conflicting Data	Number of SSNs with conflicting data	Percent of SSNs with conflicting data	Percent of all Veteran SSNs
Sex and DOB	26,345	2.2%	0.2%
Sex Only	176,346	14.7%	1.3%
DOB Only			
Year Only	118,532	9.9%	0.8%
Month Only	66,343	5.5%	0.5%
Day Only	350,577	29.2%	2.5%
Two Parts	280,853	23.4%	2.0%
Three Parts	180,049	15.0%	1.3%
Total	1,199,045	100.0%	8.6%

### 3. Concordance with Social Security Administration (SSA) and Centers for Medicare and Medicaid Services (CMS) DOB and Sex

To assess the accuracy of the VA recorded DOB, VA dates of birth were compared to dates of birth provided by SSA and CMS. CMS dates of birth are available on the VSF for Veterans enrolled in Medicare and SSA dates of birth are available on the VSF for Veterans who are deceased and have a date of death recorded on the SSA Death Master File. Only cases with a complete VA DOB and sex were used in the comparison. Table 2 contains agreement rates.

Rates of agreement between the VA DOB and the SSA DOB varied by VA data source and ranged from 70% (VBA source) to 90% (ENC source). Agreement for the year and month of birth only ranged from 85% (VBA source) to 94% (ENC source) and agreement for the year of birth only ranged from 93% (FEE source) to 96% (VBA source).

VA DOB concordance with CMS was greater when the VA sex matched CMS sex. In less than 1% of cases sex did not match. Agreement between the VA DOB and CMS DOB ranged from 75% (VBA source) to 94% (DSS source) when the sex matched, but ranged

from 26% (PTF source) to 58% (VBA source) when sex did not match. Agreement for the year and month of birth only when sex matched ranged from 87% (VBA source) to 97% (DSS source) and agreement for the year of birth only when sex matched ranged from 94% (FEE source) to 98% (DSS source).

**Table 2. Concordance of VHA DOB with SSA/CMS DOB by Source**

Source	Number of SSN/ DOB/ Sex*	DOB Match Percentage Between Source and SSA			Number of SSN/ DOB/ Sex*	DOB Match Percentage Between Source and CMS VA Sex Matches CMS Sex			Number of SSN/ DOB/ Sex*	DOB Match Percentage Between Source and CMS VA Sex Does Not Match CMS Sex		
		Full DOB	DOB Year and Month	DOB Year		Full DOB	DOB Year and Month	DOB Year		Full DOB	DOB Year and Month	DOB Year
DSS	1,763,208	89.0%	91.9%	93.2%	4,572,476	94.3%	96.8%	97.8%	11,582	39.6%	41.0%	44.7%
ENC	2,822,057	89.5%	93.3%	95.6%	6,142,285	92.1%	95.2%	97.1%	32,490	33.3%	34.6%	39.4%
ENR	2,924,743	90.4%	94.0%	95.8%	6,649,747	93.3%	96.1%	97.6%	37,924	27.4%	28.8%	34.2%
FEE	291,839	87.6%	91.5%	92.8%	393,748	89.2%	92.5%	93.7%	1,464	46.5%	49.1%	52.9%
PTF	1,739,013	88.5%	92.8%	94.7%	2,523,793	90.4%	94.1%	95.7%	9,783	25.8%	27.3%	31.8%
VBA	1,258,058	70.1%	84.7%	96.5%	2,470,961	75.0%	86.9%	97.3%	10,126	58.3%	70.5%	81.3%

\*Number of unique combinations of SSN, DOB, and Sex

Agreement varied slightly by age with generally higher rates of agreement for those under 65. Table 3 provides concordance rates for SSNs on the VSF Master File for an SSN where the first record on the file for the SSN had a complete DOB and sex and the Veteran was alive on January 1, 2010.

**Table 3. Rates of Concordance with CMS and SSA Demographics by Age Based on Age on July 1, 2011**

Parts of DOB Matching	VA Compared to CMS				VA Compared to SSA	
	VA Sex Matches CMS Sex		VA Sex Does Not Match CMS Sex			
	< 65	65 +	< 65	65 +	< 65	65 +
All Parts	97.08%	94.38%	21.63%	19.95%	96.48%	94.90%
Year and Month	98.41%	96.93%	22.75%	22.30%	97.83%	97.19%
Year	98.80%	98.69%	25.44%	31.78%	98.42%	98.12%
Number of SSNs	644,145	3,783,528	4,304	15,841	36,198	200,444

## Summary and Recommendations

In summary:

- Veteran demographics originating from most sources were generally complete (i.e., sex was recorded as male or female and DOB contained the year, month, day of birth) with the exception of the FEE source.
- Eighty-six percent (86%) of the Veterans on the VSF had a single DOB and sex combination, and an additional 5% of Veterans had multiple DOB and sex combinations only due to incomplete but not conflicting DOB or sex information (e.g., one DOB with complete information and one DOB with the day of birth missing).
- The remaining 9% of Veterans on the VSF had conflicting DOB and/or sex information.
- Agreement rates of VA year of birth with CMS year of birth and/or SSA year of birth were greater than 90% for all sources, and agreement rates with CMS dates of birth increased when the sex also matched.
- The percentage of SSNs with multiple records and the concordance rates of VA DOB with the SSA and CMS dates of birth were similar in Veterans 65 years of age or older and Veterans under 65 years of age.

When using the VHA Vital Status Master File researchers will need to plan a strategy for matching their cohort demographics to the VSF demographics that includes handling multiple demographics for the same SSN. If the VHA Vital Status Mini File, which contains only one set of demographics for an SSN, is utilized, researchers should plan to use the Master File to investigate any cases where the demographics for the SSN on the Mini File do not match the demographics from their cohort.

## References

- (1) U.S. Dept. of Veterans Affairs. Veterans Health Administration. VA Information Resource Center. "VIREC Website." Accessed Sep. 22, 2011.  
<http://www.virec.research.va.gov/>
- (2) VA Information Resource Center. *VIREC Data Investigation: VETSNET File*. Hines, IL: U.S. Dept. of Veterans Affairs, Health Services Research and Development Service, VA Information Resource Center, Jan. 2011.



**VA Information Resource Center**

Edward Hines, Jr. VA Hospital (151V)

P.O. Box 5000 • Hines, IL 60141

Ph: 708-202-2413 • Fax: 708-202-2415

[virec@va.gov](mailto:virec@va.gov) • [www.virec.research.va.gov](http://www.virec.research.va.gov)

VIREC is supported by HSR&D grant SDR 98-004.