

Creating a sampling frame for population-based veteran research: Representativeness and overlap of VA and Department of Defense databases

Donna L. Washington, MD, MPH;^{1-3*} Su Sun, MPH;¹⁻² Mark Canning, BA¹⁻²

¹Department of Veterans Affairs (VA) Greater Los Angeles Healthcare System, Los Angeles, CA; ²VA Greater Los Angeles Health Services Research and Development Center of Excellence, VA Sepulveda Ambulatory Care Center and Nursing Home, Sepulveda, CA; ³Department of Medicine, University of California Los Angeles, Los Angeles, CA

Abstract—Most veteran research is conducted in Department of Veterans Affairs (VA) healthcare settings, although most veterans obtain healthcare outside the VA. Our objective was to determine the adequacy and relative contributions of Veterans Health Administration (VHA), Veterans Benefits Administration (VBA), and Department of Defense (DOD) administrative databases for representing the U.S. veteran population, using as an example the creation of a sampling frame for the National Survey of Women Veterans. In 2008, we merged the VHA, VBA, and DOD databases. We identified the number of unique records both overall and from each database. The combined databases yielded 925,946 unique records, representing 51% of the 1,802,000 U.S. women veteran population. The DOD database included 30% of the population (with 8% overlap with other databases). The VHA enrollment database contributed an additional 20% unique women veterans (with 6% overlap with VBA databases). VBA databases contributed an additional 2% unique women veterans (beyond 10% overlap with other databases). Use of VBA and DOD databases substantially expands access to the population of veterans beyond those in VHA databases, regardless of VA use. Adoption of these additional databases would enhance the value and generalizability of a wide range of studies of both male and female veterans.

Key words: administrative databases, ambulatory care/utilization, hospital, VA nonuser, veteran indicator, veteran population, veteran sampling frame, veterans/utilization, women veterans, women's health services.

INTRODUCTION

Although the majority of veterans obtain healthcare outside the Department of Veterans Affairs (VA), most veteran research is conducted in VA healthcare settings [1–2]. While the VA databases currently in use provide

Abbreviations: AITC = Austin Information Technology Center, CP = Compensation and Pension Mini File, DEERS = Defense Enrollment Eligibility Reporting System, DMD C = Defense Manpower Data Center, DOD = Department of Defense, FY = fiscal year, HSR&D = Health Services Research and Development, NED = National Enrollment Database, NPCD = National Patient Care Database, NSV = National Survey of Veterans, NSWV = National Survey of Women Veterans, OIF/OEF = Operation Iraqi Freedom/Operation Enduring Freedom, OPC = Outpatient Care File, Q = quarter, RDD = random digit dialing, SSN = Social Security number, VA = Department of Veterans Affairs, VBA = Veterans Benefits Administration, VetPop = Veteran Population Model, VETSNET = Veterans Service Network, VHA = Veterans Health Administration, VIREC = VA Information Resource Center, VISN = Veterans Integrated Service Network.

*Address all correspondence to Donna L. Washington, MD, MPH; VA Greater Los Angeles Healthcare System, 11301 Wilshire Blvd, 111G, Los Angeles, CA 90073; 310-478-3711; fax: 310-268-4933. Email: donna.washington@va.gov

DOI:10.1682/JRRD.2009.08.0127

useful data on VA healthcare settings and characteristics of VA healthcare users, exclusion of VA nonusers leaves knowledge gaps of the health and healthcare for the vast majority of veterans. VA healthcare users and VA nonusers differ in demographic characteristics, health status, healthcare utilization patterns, and multiple other characteristics [3–5]. Ascertaining the perspective of veterans who do not use VA healthcare can provide insight on issues affecting the broader veteran community, such as barriers to healthcare use, that may be useful for VA program planning and for expanding VA access [5].

One barrier to conducting general population veteran studies is the absence of a sampling frame that includes contact information from which to identify and recruit potential research participants. Population-based registries exist for selected veteran cohorts [6–8]. However, no one database or registry exists with complete veteran population coverage. Information about population representativeness and systematic biases of available VA and non-VA administrative databases could facilitate their use for veteran research. Determining database overlap could also facilitate decision making about the marginal benefit of adding each database to a sampling frame.

Our objective was to determine the representativeness, overlap, and relative contributions of selected Veterans Health Administration (VHA), Veterans Benefits Administration (VBA), and Department of Defense (DOD) administrative databases. We conducted the National Survey of Women Veterans (NSWV) in 2008–2009 in support of VHA strategic planning for outreach, programs, and services for women veterans. Because market penetration among women veterans has been traditionally low, assessment of alternate sampling frames through use of non-VHA databases was essential. Using the NSWV sampling frame creation as an example, we report on characteristics of VHA, VBA, and DOD administrative databases.

METHODS

In 2008, we merged records from VHA, VBA, and DOD databases to identify U.S. women veterans. Databases were selected for inclusion if they included real Social Security numbers (SSNs) and data fields for contact information, sex, and veteran status. Real SSNs were important because all databases are organized by SSNs. We used real SSNs to merge databases and extracted all records of female veterans.

Source Databases

The databases we used to identify the population of women veterans were the VHA National Enrollment Database (NED) and National Patient Care Database (NPCD), the VBA Compensation and Pension Mini File (CP) and Veterans Service Network (VETS NET), and the DOD Defense Enrollment Eligibility Reporting System (DEERS).

The NED, created in 1995 and housed at the Austin Information Technology Center (AITC, previously known as the Austin Automation Center), derives data from local VA facilities on all veterans who applied for enrollment for VA healthcare [9]. The veterans represented in NED include those who (1) are currently enrolled, regardless of whether they have actually used VA healthcare; (2) applied for enrollment, even if they did not complete enrollment; or (3) were enrolled at any time from 1995 to present, even if they no longer use VA healthcare. The NED contains contact information (address and telephone number), including updates to this contact information. Other data fields include preferred VA facility. Sex is not consistently reported in this database. The NED differs from the VHA Assistant Deputy Under Secretary for Health Enrollment Monthly File in that the latter file comprises an NED data extract, along with cost, utilization, and mortality data from other sources. We used the NED in this study.

The NPCD is the VHA's centralized relational database (a data warehouse) that receives clinical encounter data from VHA clinical information systems [9–10]. The NPCD can be queried by fiscal year (FY)—e.g., FY08 = October 1, 2007–September 30, 2008—and includes all individuals visiting the VA during that time frame. It includes a veteran indicator field, added in FY06, that has the values of “yes” (for veterans), “no” (for nonveterans) or “missing” [11]. At the end of FY08 quarter (Q) 3, we obtained NED and NPCD Outpatient Care File (OPC) records from the AITC. We queried the FY06, FY07, and FY08 Q1–Q3 OPC visit data set, using the sex and veteran indicator fields to limit our sampling to female veterans [11]. All individuals with veteran status in the NPCD would be expected to have a record in the NED.

VBA databases available at the time of our database query included the CP and VETSNET data repositories of enrollees for any veteran benefit [12]. These VBA enrollees include both enrollees and nonenrollees for VA healthcare. At the time, the VBA was in the process of converting records from the CP Mini SAS data set (SAS, Inc.; Cary, North Carolina) into the new VETSNET database. We obtained CP and VETSNET SAS data sets from

the AITC and combined records to create a complete VBA data set, denoted here as CP/VETSNET.

The DOD DEERS, implemented in 1982, is a database of Active Duty and retired military servicemembers and others entitled to TRICARE benefits, a healthcare program of the DOD military health system [13]. DEERS retains records for all veterans who separated from the military after the implementation of DEERS. Veterans who were not in the Active Duty or Reserve components or receiving a DOD benefit as of 1982 will not have a record in DEERS. * We obtained DEERS data through April 2008 from the Defense Manpower Data Center (DMDC). To facilitate identifying Gulf War-era veterans with service in Operation Iraqi Freedom or Operation Enduring Freedom (OIF/OEF), we obtained separate DEERS data sets for women veterans with military separation dates on or before September 11, 2001 (9/11) versus those with separation dates after 9/11. The post-9/11 data set was used in calculations involving the post-9/11 service period, and the data sets were combined to create the complete DEERS data set.

VA-Medicare data were not available at the time of our 2008 database query. However, to facilitate future decision making about the use of these data, we estimated the upper limits of the potential contribution of VA-Medicare data to our sampling frame by imputing from the 2003 VA Information Resource Center (VIREC) VA-Medicare data merge report [14]. That report describes VA-Medicare data coverage as of 1999. In 2001, the National Survey of Veterans (NSV) was conducted on behalf of the VA to assist in VA program planning. Of the veterans covered by Medicare, 83 percent of females and 92 percent of males were aged 65 or older [15]. Therefore, we used age as a proxy for Medicare population representativeness. Women aged 65 or older in 1999 would be 74 or older, if alive, in 2008. The difference between the 1999 VIREC Medicare coverage and the sampling frame coverage of those 74 and older comprises younger Medicare-eligible veterans, those deceased between 1999 and 2008, as well as sampling frame non coverage, underscoring the potential contribution of adding VA-Medicare data.

The VA provides official estimates and projections of the veteran population using the Veteran Population Model (VetPop) [16]. VetPop2007 is the VA's latest official esti-

mate and projection of the veteran population and their characteristics from April 1, 2000, to September 30, 2036. VetPop2007 uses U.S. Census 2000 data for the 2000 population estimate; incorporates the actual number of military separations and deaths through September 30, 2006, using DMDC data; adjusts for DOD Office of the Actuary GORGO data projections of military separations after September 30, 2006; and applies additional parameters to impute mortality status for other veterans [17]. VetPop2007 includes estimates of the veteran population by sex, age, period of military service, geographic region, and race/ethnicity. Post-2006 population estimates for sex, geographic region, race/ethnicity, and mortality are included. We obtained VetPop2007 estimates for the women veteran population as of September 30, 2008, to use as the denominator in all of our database representativeness estimates. We used U.S. Census data to obtain the U.S. adult population size by sex [18].

Data Merge Procedures and Analysis

To create the sampling frame, we added a data field in each database to identify the source database. Next we merged NED, CP/VETSNET, and DEERS databases using real SSNs. Then we combined duplicate and triplicate records (as indicated by real SSNs) into one unique record that retained the source database identifiers. We derived population representativeness of the sampling frame by dividing the number of unique records by the September 30, 2008, VetPop2007 estimate of the women veteran population. The frequency of the source database identifiers gave the size of constituent segments of the sampling frame. To determine the relative contribution of each database to the population representativeness, we first determined which database had the greatest number of records. Starting with the largest database, we identified the constituent sampling frame segments that were not a part of that largest database. We repeated this process for the second largest database. Then, using the database and constituent segment sizes as numerators, we divided each by the September 30, 2008, VetPop2007 population estimate. We also determined representativeness and overlap of each constituent database by dividing each database and segment size, respectively, by the September 30, 2008, VetPop2007 population estimate.

The NPCD records queried for the 3 FYs were merged with the use of real SSNs. Duplicate records, defined as records with identical real SSNs, were eliminated. To determine the positive predictive value of the

*Personal communication, George Fitzgerald, III, Defense Manpower Data Center (DMDC); May 27, 2008.

NPCD veteran indicator, we calculated the rate of self-reported service in the U.S. military in the sample with a positive veteran indicator that was screened for the NSWV. Those with a negative veteran indicator were not screened, and therefore other test characteristics of the veteran indicator could not be assessed. To determine the percentage of VA enrollees using VA healthcare over different time frames, we used the NED as the denominator for each of the individual NPCD queries and for the 3 NPCD years combined. To determine the sampling frame representativeness of veterans who were not current VA healthcare users, we divided the number of unique records in the combined databases by the VetPop 2007 women veteran population estimate after each was reduced by the NPCD FY08 count.

Characteristics ascertained for each database were percentage of current VA healthcare users, percentage of women veterans aged 65 or older, and percentage of women veterans with post-9/11 service who have a record in the database (from DEERS). Current VA healthcare use was defined as the last VA clinical encounter in FY08 (from NPCD). All databases contained fields for date of birth and/or age that we used to classify women veterans as being 65 years or older versus not and as 74 years or older for the imputation of potential VA-Medicare database representativeness. The number of women with each characteristic of interest in each database was used as the numerator. We conducted chi-square tests comparing women veterans in each database with those not in that database for each characteristic.

RESULTS

VetPop2007 estimates a 2008 U.S. women veteran population size of 1,802,000. VHA, VBA, and DOD databases combined yielded 925,946 unique records, representing 51.4 percent of U.S. women veterans. The relative contribution of each database to the population representativeness is illustrated in the **Figure**, which depicts the size and overlap of each circle in relative proportion to the database contribution to the combined sampling frame. The DEERS database contributed the greatest number of records, representing 29.7 percent of the U.S. women veteran population. In the combined sampling frame, 42 percent of all records, and 124,938 of the 167,117 (74.8%) separated after 9/11, were solely in DEERS (DEERS only = segment A). Combining NED

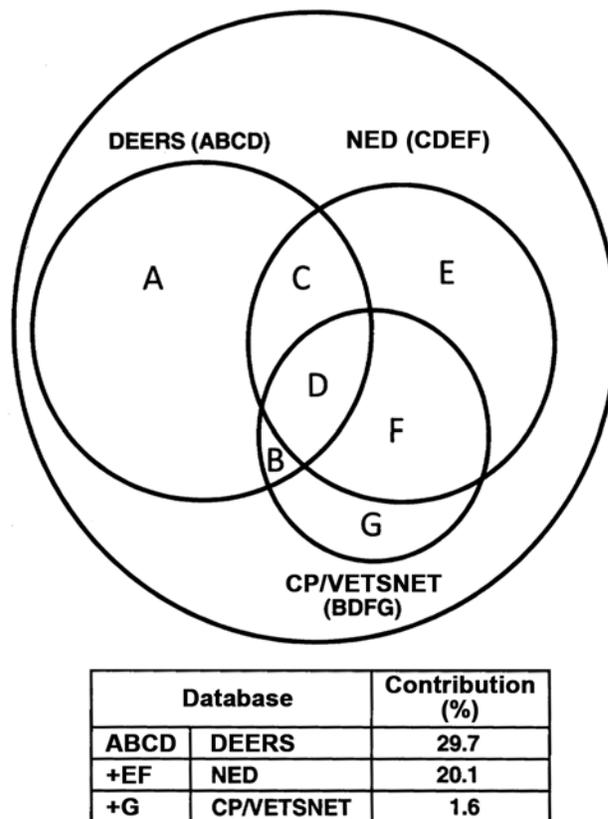


Figure.

Relative contributions of selected Department of Defense, Veterans Health Administration, and Veterans Benefits Administration databases for representing 2008 U.S. women veteran population ($N = 1,802,000$). **Table 1** in main text shows size of each segment, A–G. DEERS = Defense Enrollment Eligibility Reporting System, NED = National Enrollment Database, CP/VETSNET = Compensation and Pension Mini File and Veterans Services Network (combined).

with DEERS extends the sampling frame representativeness to an additional 20.1 percent of the U.S. women veteran population, for a population representation of 49.8 percent from these two databases alone.

The size and percentage population representativeness of each database and its overlap with other databases are given in **Table 1**. A majority of DEERS and of NED records were solely in those databases (72.4% and 50.9%, respectively). Although CP/VETSNET included 11.8 percent of the women veteran population, only 13.9 percent of its records were solely in that database.

The NPCD veteran indicator field had a positive predictive value of 99.6 percent for correctly identifying veterans. Of VA enrollees with a positive veteran indicator, the

Table 1.

Representativeness and overlap of selected Veterans Health Administration (VHA), Veterans Benefits Administration (VBA), and Department of Defense (DOD) databases: 2008 women veteran population.*

Database	Women Veterans (n)	U.S. Women Veteran Population (%)
DEERS	534,366	29.7
A. DEERS only	386,873	21.5
B. DEERS + CP/VETSNET	8,901	0.5
C. DEERS + NED	71,418	4.0
D. DEERS + NED + CP/VETSNET	67,174	3.7
NED	500,651	27.8
E. NED only	254,622	14.1
F. NED + CP/VETSNET	107,437	6.0
CP and VETSNET	213,033	11.8
G. CP/VETSNET only	29,521	1.6
Combined Databases	925,946	51.4

*Denominator is 1,802,000 for Veteran Population Model 2007 estimates and projections for September 30, 2008, women veteran population from National Center for Veterans Analysis Statistics.

DEERS = Defense Enrollment Eligibility Reporting System (DOD), NED = National Enrollment Database (VHA), CP/VETSNET = Compensation and Pension Mini File (VBA) and Veterans Service Network (combined databases).

number and percentage with current and recent VA healthcare use are given in **Table 2**. Among women veteran VA enrollees, 50.6 percent used VA healthcare in the prior 9 months. An additional 9.7 percent of enrollees last used VA more than 9 but less than 33 months prior. In the VA enrollee database, 39.7 percent of records were for women veterans who either last used VA healthcare more than 33 months prior or had never used VA healthcare. In the overall U.S. women veteran population, 14.1 percent were current VA healthcare users (253,314 NPCD FY08 Q1–Q3 women veterans out of the 1,802,000 U.S. women veteran population; **Table 2**), whereas 72.2 percent were not enrolled for VA care (derived from 500,651 NED enrollees comprising 27.8 percent of the U.S. women veteran population in **Table 2**). Approximately 673,000 sampling frame records were for women veterans who were not current VA healthcare users, covering 43.4 percent of the U.S. women veteran VA nonuser population.

Characteristics of each of the databases are listed in **Table 3**. By definition, all current VA healthcare users had records in the NED, whereas only 27.3 percent had records in DEERS ($p < 0.001$ for comparisons among all three databases). All databases contained records for only a minority of women veterans age 65 or older, with CP/VETSNET and DEERS having significantly smaller pro-

Table 2.

Number and percentage of Department of Veterans Affairs (VA) women veteran enrollees with current and recent VA healthcare use.

Database	Women Veterans Enrollees (n)	Women Veteran Enrollees (%)	U.S. Women Veterans (%)*
NPCD FY08 Q1–Q3	253,314	50.6	14.1
NPCD FY07	245,393	49.0	13.6
NPCD FY06	231,869	46.3	12.9
NPCD FY06–FY08 Q3	301,951	60.3	16.8
NED	500,651	100.0	27.8

*Denominator is 1,802,000 from Veteran Population Model 2007 estimates and projections for September 30, 2008, women veteran population from National Center for Veterans Analysis and Statistics.

FY = fiscal year (October 1–September 30), NED = National Enrollment Database (Veterans Health Administration [VHA]), NPCD = National Patient Care Database in VHA for VA healthcare users during indicated FYs, Q = quarter.

portions than NED ($p < 0.001$). The VA-Medicare database included 100,329 female enrollees as of 1999, whereas the combined sampling frame included approximately 65,000 aged 74 or older as of 2008. This difference of approximately 35,000 is 1.9 percent of the U.S. women veteran population. The percentage of women veterans with post-9/11 service also varied significantly by database, with representativeness being smallest for the CP/VETSNET database ($p < 0.001$).

DISCUSSION

With a combination of VHA, VBA, and DOD administrative databases, we assembled a sampling frame that contains more than half of living U.S. women veterans. Sampling frame representativeness varied by U.S. women veteran subpopulation. The sampling frame contained records for all VA healthcare users and for all gulf war-era veterans, including OIF/OEF women veterans who were not on Active Duty status. Forty-three percent of VA nonusers were included in the database, with the smallest proportionate representation being among veterans aged 65 and older—a group that is likely to need healthcare services for chronic medical conditions. In prior research, we found that some older women veterans briefly used VA healthcare in the remote past and then never returned, which could contribute to the underrepresentation of older women veterans in VHA databases [19]. Many older women veterans separated from the military prior to the creation of DEERS,

Table 3.

Characteristics of selected Veterans Health Administration (VHA), Veterans Benefits Administration (VBA), and Department of Defense (DOD) databases: 2008 women veteran population.

Characteristic*	NED (<i>n</i> = 500,651) (%)	CP/VETSNET (<i>n</i> = 213,033) (%)	DEERS (<i>n</i> = 534,366) (%)
Current VA Users (<i>n</i> = 253,314) [†]	100.0	44.9	27.3
Age 65 or Older (<i>n</i> = 309,000) [‡]	27.6	5.5	1.5
Post-9/11 Service (<i>n</i> = 167,117; includes OIF/OEF women veterans) [‡]	23.5	13.7	100.0

**p* < 0.001 for comparisons among databases for each characteristic.

[†]Fiscal year 2008, quarters 1–3.

[‡]Population estimates for September 30, 2008, Veteran Population Model rounded to nearest thousand from National Center for Veterans Analysis and Statistics.

9/11 = September 11, 2001; CP/VETSNET = Compensation and Pension Mini File (VBA) and Veterans Service Network (combined databases); DEERS = Defense Enrollment Eligibility Reporting System (DOD); NED = National Enrollment Database (VHA); OIF/OEF = Operation Iraqi Freedom/Operation Enduring Freedom; VA = Department of Veterans Affairs.

which could account for their underrepresentation in the DOD database. The use of Medicare data could provide a more focused sampling frame of older individuals to screen for veteran studies. This could potentially extend veteran identification to an additional small percentage of women veterans and likely a higher proportion of male veterans, given their older age.

One approach to using a sampling frame in which differential representation of segments of the target population has been defined is to adjust for this systematic bias with the use of stratified sampling and/or the application of weights. For example, in a prior study that utilized a similar approach to assemble a sampling frame of women veterans residing in southern California and southern Nevada (corresponding to Veterans Integrated Service Network [VISN] 22), we applied separate sampling strategies for VA users and VA nonusers and adjusted for the design effect and nonresponse in our sampling weights [4]. Because 100 percent of VA users were represented in the database, for each user strata, we randomly selected VA users from each of the corresponding age categories. Because the source databases differed in size and their VA user/nonuser ratios, for each VA nonuser strata, we assigned sampling probabilities for subjects based on their source database that correlated with how that database represented the VA user/nonuser distribution in the VISN 22 women veteran population. Our rationale for this approach was to minimize the effects of potential systematic biases in the databases that may be associated with decision making about VA healthcare use. This approach may be less biased than a sample drawn randomly from among VA nonusers in our source databases. For studies in which the sampling frame contains all members of the target population—as would be

the case for population-based studies of veterans who separated from the military in the past 20 years or less—a simple random sample of the appropriate subset of the sampling frame may be used.

Although stratified sampling and weighting can reduce the underrepresentation bias associated with sampling from administrative databases (i.e., list-based sampling), they are unlikely to completely eliminate bias. Therefore, some surveys use random digit dialing (RDD) to identify potential respondents. For example, the NSV combines a list-based approach for identifying smaller subpopulations of veterans who are targeted for oversampling (selected VA priority groups, ethnic minorities, and women in the 2001 survey) with RDD for identifying larger subpopulations of veterans [1–5]. The NSV employed this dual-frame approach to combine the efficiency and low cost of list-based sampling with the greater population representativeness of RDD. The 2001 NSV did not rely solely on a list-based approach because it was only able to identify 21.6 percent of the overall veteran population using administrative databases; by contrast, we achieved 51.4 percent population identification. The 2001 NSV did not rely solely on an RDD approach because it estimated that it would have cost more than five times the dual-frame approach [15]. An additional limitation of RDD is underrepresentation of those with no telephone or an unlisted number, which is estimated to be 4 to 6 percent [20]. For the 1985 Survey of Female Veterans, one of the few large veteran surveys to rely solely on an RDD approach, over 880,000 numbers were dialed to reach 3,925 women veterans [21]. By contrast, 19 percent of U.S. adult males are veterans, making RDD a useful approach for general population studies of male veterans. For smaller subpopulations of veterans,

such as women veterans (comprising 1.5% of the U.S. adult female population), or those with other selected conditions or characteristics of interest, a solely RDD approach is not a viable alternative.

Most administrative databases were not designed with health services research use as a primary goal, and therefore they are limited in their use in research. Using only exact SSN matches to merge databases gets the most likely matches but misses those who have SSN errors in one or more databases. The NED misses those who never enrolled for VA care (nonenrollees) and those who have not enrolled for VA care since 1995 (remote users). Users of veteran centers, community-based counseling centers that provide readjustment counseling and outreach services to combat veterans and their family members, would not have an NED record if they were VA nonenrollees or remote users. NED data fields for demographic variables such as race/ethnicity and sex may also be inaccurate or incomplete. For example, our sampling frame did not include for consideration the very small number of records in which sex was not recorded. Some care is delivered to veterans by non-VA providers under contract to the VA. This includes extended care at community nursing homes and state veterans homes and specialized women's health services [22]. Although veterans using this care will have an NED record, the NPCD misses most contract care and therefore veterans using only contract care could be misclassified as last using the VA more than 3 years ago. The DEERS database does not systematically update the contact information in their records. The characteristics associated with accuracy of this information should be investigated in the future. The DEERS database also does not cover veterans who were not in the Active Duty or Reserve components or who were not receiving a DOD benefit as of 1982. Given the younger age of women veterans compared with male veterans, our findings may underestimate the proportionate Medicare database representativeness and overestimate the DEERS representativeness for male veterans. Finally, the accuracy of the VetPop2007 model is unknown. Population estimates after 2006 are based on projections rather than actual data and therefore may have decreasing accuracy for increasing time frames beyond 2006. Although this method of estimation might have a small effect on our 2008 sampling frame coverage estimates, this would not alter the relative contributions of the different databases to the overall sampling frame.

CONCLUSIONS

The contributions of our approach to veteran sampling frame creation are its expansion of list-based veteran population representativeness beyond that reported by other researchers, its elucidation of the relative contributions of various databases and database characteristics to facilitate decision making about their use, and its characterization of the NPCD veteran indicator field. Combining VHA, VBA, and DOD administrative databases substantially expands access to the population of veterans beyond VA healthcare users. Adoption of these additional databases would enhance the value and generalizability of a wide range of studies of both male and female veterans. Using our approach to assembling VA and DOD administrative databases, several subpopulations of veterans may be ascertained with minimal bias. These include VA healthcare users, VA enrollees, and veterans who separated from the military in the prior 20 years, including gulf war-era veterans and OIF/OEF veterans. Veterans and those persons engaged in clinical practice, policy, and research involving veterans may benefit from findings generated from research studies that include the perspective of both VA users and veterans who do not use VA healthcare.

ACKNOWLEDGMENTS

Author Contributions:

Study concept and design: D. L. Washington.

Acquisition of data: M. Canning.

Analysis and interpretation of data: D. L. Washington, S. Sun.

Drafting of manuscript: D. L. Washington.

Critical revision of manuscript for important intellectual content:

D. L. Washington, S. Sun, M. Canning.

Statistical analysis: S. Sun.

Obtaining funding: D. L. Washington.

Administrative and technical support: M. Canning.

Study supervision: D. L. Washington.

Financial Disclosures: The authors have declared that no competing interests exist.

Funding/Support: This material was based on work supported in part by the VA, Office of Public Health and Environmental Hazards, Women Veterans Health Strategic Healthcare Group (NSWV), with supplemental funding from the VA Health Services Research and Development Service (HSR&D) (SDR-08-270).

Additional Contributions: We thank James Strike, Robert Thomas-Cano, and Michael Wasson of the AITC; Jane Styer, George Fitzgerald III, and Scott Seggerman of the DMDC; and Mingming Wang of the VA Greater Los Angeles HSR&D Center of Excellence for assistance

with sample development. The views expressed within are solely those of the authors, and do not necessarily represent the views of the VA.

Institutional Review: This research was conducted as part of the NSWV that was approved by the Institutional Review Board and the Research and Development Committee of the VA Greater Los Angeles Healthcare System and by the U.S. Office of Management and Budget.

REFERENCES

1. Goldzweig CL, Balekian TM, Rolón C, Yano EM, Shekelle PG. The state of women veterans' health research. Results of a systematic literature review. *J Gen Intern Med.* 2006; 21(3):S82–92. [PMID: 16637952] [DOI:10.1111/j.1525-1497.2006.00380.x](https://doi.org/10.1111/j.1525-1497.2006.00380.x)
2. Weeks WB, Wallace AE, West AN, Heady HR, Hawthorne K. Research on rural veterans: An analysis of the literature. *J Rural Health.* 2008;24(4):337–44. [PMID: 19007387] [DOI:10.1111/j.1748-0361.2008.00179.x](https://doi.org/10.1111/j.1748-0361.2008.00179.x)
3. Washington DL, Villa V, Brown A, Damron-Rodriguez J, Harada N. Racial/ethnic variations in veterans' ambulatory care use. *Am J Public Health.* 2005;95(12):2231–37. [PMID: 16257951] [DOI:10.2105/AJPH.2004.043570](https://doi.org/10.2105/AJPH.2004.043570)
4. Washington DL, Yano EM, Simon B, Sun S. To use or not to use—What influences why women veterans choose VA healthcare. *J Gen Intern Med.* 2006;21(Suppl 3):S11–18. [PMID: 16637939] [DOI:10.1111/j.1525-1497.2006.00369.x](https://doi.org/10.1111/j.1525-1497.2006.00369.x)
5. Long JA, Polsky D, Metlay JP. Changes in veterans' use of outpatient care from 1992 to 2000. *Am J Public Health.* 2005;95(12):2246–51. [PMID: 16257943] [DOI:10.2105/AJPH.2004.061127](https://doi.org/10.2105/AJPH.2004.061127)
6. Henderson WG, Eisen S, Goldberg J, True WR, Barnes JE, Vitek ME. The Vietnam Era Twin Registry: A resource for medical research. *Public Health Rep.* 1990;105(4):368–73. [PMID: 2116638] [DOI:10.1111/j.1525-1497.2006.00369.x](https://doi.org/10.1111/j.1525-1497.2006.00369.x)
7. Murphy FM, Kang H, Dalager NA, Lee KY, Allen RE, Mather SH, Kizer KW. The health status of Gulf War veterans: Lessons learned from the Department of Veterans Affairs Health Registry. *Mil Med.* 1999;164(5):327–31. [PMID: 10332170] [DOI:10.1111/j.1525-1497.2006.00369.x](https://doi.org/10.1111/j.1525-1497.2006.00369.x)
8. Rabeneck L, Menke T, Simberkoff MS, Hartigan PM, Dickinson GM, Jensen PC, George WL, Goetz MB, Wray NP. Using the national registry of HIV-infected veterans in research: Lessons for the development of disease registries. *J Clin Epidemiol.* 2001;54(12):1195–1203. [PMID: 11750188] [DOI:10.1016/S0895-4356\(01\)00397-3](https://doi.org/10.1016/S0895-4356(01)00397-3)
9. VHA corporate databases monograph. 2007 [Internet]. Washington (DC): VHA Office of Information Health Informatics and Information Resources National Data Systems (NDS); 2007 [cited 2009 Jul 31]. Available from: <http://www.virec.research.va.gov/References/Links/VHACorporateDatabaseMonograph2007.pdf>
10. National Patient Care Database (NPCD). Washington (DC): VA Information Resource Center; 2010 [updated 2010 Jan 13; cited 2009 Jul 31]. Available from: <http://www.virec.research.va.gov/DataSourcesName/NPCD/NPCD.htm>
11. King S, Shane A, Smith MW. A guide to identifying non-veteran records in the inpatient and outpatient databases [Internet]. Menlo Park (CA): HERC; 2006. Technical Report #20. Available from: http://www.herc.research.va.gov/files/RPRT_414.pdf
12. Compensation and Pension /VETSNET data [Internet]. Washington (DC): National Data Systems; 2009 [updated 2009 Sep 11; cited 2009 Jul 31]. Available from: <http://vaww.va.gov/NDS/CompensationAndPension-VETSNETData.asp>
13. DEERS overview [Internet]. Washington (DC): Defense Enrollment Eligibility Reporting Systems; 2009 [cited 2009 Jul 31]. Available from: <https://www.dmdc.osd.mil/appj/deerswebsite/home.do;jsessionid=a462e9e5a0712aa833fdad25e2c84cad0efc3848d25912610339e56f8bc98ae0>
14. Research findings from the VA-Medicare data merge initiative: Veterans' enrollment, access and use of Medicare and VA Health Services (X VA 69-001) [Internet]. Hines (IL): Department of Veterans Affairs; 2003 [cited 2009 Jul 31]. Available from: <http://www.virec.research.va.gov/DataSourcesName/VA-CMS/Medicare/USHreport.pdf>
15. Choudhry GH, Park I, Kudela MS, Helmick JC. 2001 National Survey of Veterans Design and Methodology [Internet]. Washington (DC): Department of Veterans Affairs; 2002 [cited 2009 Jul 31]. Available from: http://www1.va.gov/VETDATA/docs/SurveysAndStudies/NSV_Methodology_Report.pdf
16. Veteran Population 2007 [Internet]. Washington (DC): Department of Veterans Affairs; 2007 [updated 2010 Mar 22; cited 2009 Jul 31]. Available from: <http://www1.va.gov/VETDATA/Demographics/Demographics.asp>
17. Veteran Population Model [Internet]. Washington (DC): Department of Veterans Affairs; 2007 [cited 2009 Jul 31]. Available from: <http://www.nhpco.org/files/public/veterans/VA%20101%20Documents/VetPopModel.pdf>
18. DP-1. General demographic characteristics [Internet]. Washington (DC): U.S. Census Bureau; 2008 [cited 2009 Jul 31]. Available from: http://factfinder.census.gov/servlet/QTTable?_bm=y&-qr_name=PEP_2008_EST_DP1&-geo_id=01000US&-ds_name=PEP_2008_EST&-lang=en&-format=&-CONTEXT=qt

19. Washington DL, Kleimann S, Michelini AN, Kleimann KM, Canning M. Women veterans' perceptions and decision-making about Veterans Affairs health care. *Mil Med.* 2007; 172(8):812–17. [\[PMID: 17803071\]](#)
20. Brick JM, Waksburg J, Kulp D, Starer A. Bias in list-assisted telephone samples. *Public Opin Q.* 1995;59:218–35. [DOI:10.1086/269470](#)
21. Survey of female veterans. Washington (DC): Department of Veteran Affairs; 1985.
22. Washington DL, Caffrey C, Goldzweig C, Simon B, Yano EM. Availability of comprehensive women's health care through Department of Veterans Affairs medical centers. *Women's Health Issues.* 2003;13(2):50–54. [DOI:10.1016/S1049-3867\(02\)00195-0](#)

Submitted for publication August 14, 2009. Accepted in revised form May 6, 2010.

This article and any supplementary material should be cited as follows:

Washington DL, Sun S, Canning M. Creating a sampling frame for population-based veteran research: Representativeness and overlap of VA and Department of Defense databases. *J Rehabil Res Dev.* 2010;47(8):763–72.

DOI:10.1682/JRRD.2009.08.0127



