Sourcebook:
Women Veterans in the Veterans Health Administration

Volume 4: Longitudinal Trends in Sociodemographics, Utilization, Health Profile, and Geographic Distribution

Online Appendix
(Technical Appendix)

Recommended citation: Frayne SM, Phibbs CS, Saechao F, Friedman SA, Shaw JG, Romodan Y, Berg E, Lee J, Ananth L, Iqbal S, Hayes PM, Haskell S. Sourcebook: Women Veterans in the Veterans Health Administration. Volume 4: Longitudinal Trends in Sociodemographics, Utilization, Health Profile, and Geographic Distribution. Women's Health Evaluation Initiative, Women's Health Services, Veterans Health Administration, Department of Veterans Affairs, Washington DC. February 2018.

Table of Contents

List of Exhibits

1.0 Data Sources

Data for Sourcebook Volume 4 came from centralized VHA administrative data files. The source files used to create the Sourcebook database are:[1]

ADUSH Fiscal Year-End Enrollment File (FY00-FY15). These VHA enrollment data files are maintained by the Office of the Assistant Deputy Under Secretary for Health (ADUSH) and contain records of patient characteristics (sex, Veteran status, VHA user status, date of birth, service-connected (SC) disability status, etc.). Enrollment files used span a 16-year period from fiscal year (FY) 2000 through fiscal year 2015.[2] Sourcebook Volume 4 refers to this as the "ADUSH Enrollment File."

VHA Medical SAS Datasets

a. **VHA Outpatient Event and Visit Files (Medical SAS Outpatient Datasets, FY00–FY15).** The Outpatient Event (SE) file contains a record for every encounter the patient has with VHA (e.g., clinic visits, telephone encounters, lab test encounters, radiology encounters, etc.); there can be more than one encounter on a given day. Each record contains information about the encounter (e.g., date of care, VHA facility, clinic types, diagnoses associated with the visit, procedures performed at the visit, etc.). The Outpatient Visit (SF) file consolidates records of SE file encounters into one record per day of care, and provides additional information about patients (e.g., sex, date of birth, etc.).
   i. MDPPRD.MDP.SAS.SEyy (SE)
   ii. MDPPRD.MDP.SAS.SFyy (SF)

b. **VHA Inpatient Main and Bed Section Files (Medical SAS Inpatient Datasets, FY00–FY15).** These VHA inpatient stay files contain a record for every admission to a VHA facility. This includes admissions to acute care settings (e.g., medical/surgical, psychiatric, etc.), observation bed stays, and extended care stays. The inpatient stay files include patient demographic data as well as information on diagnoses, procedures, and surgeries performed while an inpatient.
   i. MDPPRD.MDP.SAS.PMyy
   ii. MDPPRD.MDP.SAS.PMOyy
   iii. MDPPRD.MDP.SAS.XMyy
   iv. MDPPRD.MDP.SAS.CENSUS.PMyy
   v. MDPPRD.MDP.SAS.CENSUS.PMOyy
   vi. MDPPRD.MDP.SAS.CENSUS.XMyy
   vii. MDPPRD.MDP.SAS.PByy
   viii. MDPPRD.MDP.SAS.PBOyy
   ix. MDPPRD.MDP.SAS.XByy
   x. MDPPRD.MDP.SAS.CENSUS.PByy
   xi. MDPPRD.MDP.SAS.CENSUS.PBOyy
   xii. MDPPRD.MDP.SAS.CENSUS.XByy

This Sourcebook refers to files i-vi, above (the Medical SAS Inpatient Main files) as the "Inpatient Main files" and files vii-xii, above (the Medical SAS Inpatient Bed Section files) as "Inpatient Bed Section files." The Inpatient Bed Section files are used in this Sourcebook only as part of the Health Profile

[1] Detailed descriptions of source datasets can be found on the VA Information Resource Center website (available at http://vaww.virec.research.va.gov/Intro/Working-with-VA-Data.htm) and the VA Health Economics Resource Center website (available at: http://www.herc.research.va.gov/).
[2] FY15 is October 1, 2014, through September 30, 2015.

identification algorithm (see Section 6.0).

VHA Vital Status File (FY16Q3). The VHA Vital Status File (VSF) contains mortality and demographic data, including race and ethnicity data from Medicare, for all Veterans who are enrolled in VHA, who received VHA care since 1992, or who have received Veterans Benefits Association (VBA) compensation or pension benefits since 2002.

VA OEF/OIF/OND Roster (FY15Q2). The DOD, Manpower Data Center (DMDC) maintains a cumulative roster of all Veterans who have been deployed to the Operation Enduring Freedom/Operation Iraqi Freedom/Operation New Dawn (OEF/OIF/OND) missions or those who have served in support of the missions, and whose most recent military discharge occurred after October 1, 2001. VA Post Deployment Health Services Epidemiology Program maintains the VA OEF/OIF/OND Roster. The Roster includes the subset of these Veterans who were enrolled in VHA prior to the date the VA OEF/OIF/OND Roster was updated—we used the version updated in January 2015, the most recent version available as of the time of this work.

Managerial Cost Accounting (MCA)[3] National Data Extracts (NDEs) (FY00-FY15). The MCA is the cost accounting system for the Department of Veterans Affairs. The NDEs are a set of data files containing the costs of VHA-provided inpatient and outpatient encounters.

a. **OPAT.** The MCA OPAT contains costs for VHA outpatient clinic encounters, as well as outpatient laboratory, pharmacy, ancillary services, prosthetics, and other services. Costs for care provided by some non-VA long term care facilities (such as state Veterans homes) are also included.
b. **Pharmacy File.** The MCA Pharmacy File includes information and costs of drugs used by all patients in both the outpatient and inpatient settings.

Purchased Care (Fee) Files (FY00-FY15). Care provided in Non-VA facilities but paid for by VA is recorded in the Purchased Care files. The file for a given fiscal year contains a record for each service reimbursed in that fiscal year, along with other information (e.g., date of service, type of service, diagnoses associated with the service, amount paid to the service provider, etc.). Payments made in a given fiscal year may reflect services provided in that fiscal year or in previous years.[4]

a. **Purchased Care Outpatient Services File.** Called "MDPPRD.MDP.SAS.FEN.FYyy.MED," the Purchased Care Outpatient Services file reflects services provided through the Purchased Care system. It includes services provided by non-VHA providers in the fiscal year reflected in the file name, or services provided in prior years that VHA reimbursed in the fiscal year. For example, the FY15 Purchased Care outpatient file contains many records reflecting services provided during FY15, but also contains some records reflecting services provided during prior years.
b. **Purchased Care Inpatient Stay and Ancillary Files.** The Purchased Care Inpatient Stay file contains a record for each submitted invoice not exceeding the allowable Medicare Diagnostic Related Group (DRG) payment. The Inpatient Ancillary file contains records for services whose reimbursement exceeds the Medicare DRG amount as well as for physician care provided in the inpatient setting.
    i. MDPPRD.MDP.SAS.FEN.FYyy.INPT
    ii. MDPPRD.MDP.SAS.FEN.FYyy.INPT.ANCIL

[3] Formerly Decision Support System or DSS.

[4] Therefore, some records contained in Purchased Care files reflect care provided in the prior fiscal year or earlier. Conversely, some care provided in a given fiscal year will not appear in that fiscal year's Purchased Care file but instead will appear in a subsequent year's file.

Non-VA Inpatient Stays Files (FY00-FY15). The Non-VA Inpatient File contains additional data on non-VA inpatient stays entered into the Medical SAS Inpatient Datasets.

VHA Site Tracking (VAST) Database Quarterly Executive Summary (FY15Q4). The VAST site list, maintained by VHA Support Service Center (VSSC), is the official VHA list of all sites of care and includes the mappings of the individual sites to parent station and VISN. Quarterly updates are available – we used the FY15 Quarter 4 version.

All programming was performed using SAS 9.2©, and all programs were independently validated by at least one other data analyst. Data presented in this Sourcebook were analyzed for program evaluation purposes.[5]

[5] These program evaluation analyses are for non-research purposes.

2.0 Cohort Creation

2.1. Overview

We first describe construction of the WHEI Master Database for each VA fiscal year, which includes one record for each person who appears in the ADUSH Enrollment File for that fiscal year. We then describe how, for Sourcebook Volume 4, we selected Veteran VHA users in each fiscal year from the WHEI Master Database to create the Base Cohort of Veteran VHA patients for each fiscal year (FY00, FY05, FY10, and FY15).

2.2. WHEI Master Database

We created person-level analytical files with one observation for each person found in the ADUSH Enrollment File. The resulting WHEI Master Database includes the following types of people:

- Veterans and non-Veterans
- Users and non-users of VHA care
- Women and men

Each is identified in the WHEI Master Database by scrambled Social Security Number. Various record-level files described above in Section 1.0 were used to create multiple person-level variables for each individual; variable specifications are described in subsequent sections. Year-specific variables indicate whether an individual was a VHA user or a Veteran in a given year, since these are characteristics that may change over time.

2.3. Base Cohort for Sourcebook Volume 4

To create the Base Cohort for each fiscal year (FY00, FY05, FY10, and FY15), we selected all women and men Veterans[6] who, based on the ADUSH Enrollment File, used VHA for outpatient and/or inpatient care and/or Purchased Care services and/or non-VA contract care and/or pharmacy services at least once in the years being examined, from FY00–FY15.[7]

For each year examined, Exhibit A shows the number of people in the WHEI Master Database, and the number in the Base Cohort (i.e., the number within the WHEI Master Database who were Veteran VHA users).[8]

[6] Non-Veterans who use VHA services are not included in the current Sourcebook. Previous work (Frayne SM., Yano EM, Nguyen VQ, et al. Gender disparities in Veterans Health Administration care: Importance of accounting for veteran status. *Med Care* 2008;46 (5):549-553) has found that, in FY02, nearly half the women in the SE data files were non-Veterans, and the majority of these non-Veterans were employees. Employees appear in the database primarily due to their encounters with Employee Health (e.g., for mandatory tuberculosis screening or for influenza vaccines). Other non-Veterans who use VHA services include some active duty military personnel and some eligible spouses of Veterans. They are not a focus of Sourcebook Volume 4 but represent a subgroup of women seen in primary care and other settings.

[7] Because the ADUSH Enrollment File counts use of non-VA contract care and pharmacy services as instances of VHA utilization, a small number of patients whose only use of VHA services is through non-VA contract care or outpatient pharmacy services are included in the cohort examined in Sourcebook Volume 4. In FY15, only 1.7% (5,420) of women Veterans who were identified as VHA users by ADUSH Enrollment Files had no utilization of VHA outpatient or inpatient care or Purchased Care (and thus were presumably VHA users by virtue of non-VA contract care or pharmacy services only). Because the analyses in Sourcebook Volume 4 do not draw upon VHA's contract care or pharmacy files, we do not explicitly characterize these types of utilization, even though those patients are included in our total counts of VHA patients.

[8] The definitions of variables for Veteran, VHA user and sex are intended to replicate as closely as possible the definitions used by the VHA Support Service Center (VSSC) in their data report cubes, so as to maximize compatibility between data appearing in various VHA reports.

Exhibit A. Number of Individuals in WHEI Master Database and in Base Cohort by Year

Fiscal Year	WHEI Master Database: Total Number (Veterans and non-Veterans, VHA users and non-users, women and men)	Base Cohort: Total Number (Veterans only, VHA users only, women and men)*
2000	5,249,069	3,395,205
2005	8,206,947	4,802,582
2010	9,216,578	5,351,873†
2015	10,170,824	5,891,311

* These numbers include those with missing sex values. For example, in FY15, there were 5,891,311 Veteran VHA users in the WHEI Master Database. This is 1,237 more people than the 5,890,074 women plus men Veteran users reported in Sourcebook Volume 4, because of missing sex data. Sex was not available for 1,237 Veteran users in FY15. See Section 3.3 for the number of patients in each year with missing sex data.

† Sourcebook Volume 2 reports on Veteran VHA patients in FY10. The number of Veteran VHA patients in FY10 reported in this Sourcebook Volume 4 (N=5,351,873) differs slightly from the number of Veteran VHA patients in FY10 reported in Sourcebook Volume 2 (N= 5,354,652). This is due to changes in the WHEI Veteran algorithm. As of FY10, ADUSH Enrollment Files no longer use the "ELIG" variable to determine Veteran status, and so WHEI has modified the Sourcebook Volume 4 Veteran status algorithm to be consistent with the ADUSH Enrollment Files' Veteran status algorithm. Due to these updates and further changes in algorithms used to define sex and date of birth, the number of women Veteran patients in FY10 reported in this Sourcebook volume (N=317,122) differs from the 316,903 women Veteran patients reported in Volume 2 and the 316,414 women Veteran patients reported in Volume 3. See Section 3.0 for descriptions of these algorithms.

3.0 Algorithms for Part 1. Sociodemographics

The Base Cohort created for the Sourcebook series includes person-level sociodemographic variables derived from data in the ADUSH Enrollment File (in some cases supplemented with data from the VHA Medical SAS Datasets, MCA NDEs, Non-VA Inpatient Files, VHA VSF, and VA OEF/OIF/OND Roster)[9] for each year from FY00–FY15. These variables include Veteran status, sex, date of birth, race/ethnicity, urban/rural status, and SC disability rating status. For all algorithms, we describe the steps taken to create the FY15 variable, unless otherwise specified.

3.1. Veterans

Three methods were employed to identify Veterans (yes/no variable) in the ADUSH Enrollment Files in each year over the 16-year period (FY00-FY15). Because true changes in Veteran status can occur, we did not require that Veteran status for a given individual be consistent across years.

FY00-FY02: Identified using fields labeled "MATCH" and "ELIG"

A patient is considered a Veteran if both of the following are true:
1. MATCH value is NOT "COSTONLY", AND
2. The first letter of the ELIG field value is NOT="N".

FY03–FY09: Identified using fields labeled "PRIO1_8" and "ELIG".

A patient is considered a Veteran if either of the following is true:
1. PRIO1_8 value is NOT missing OR
2. PRIO1_8 value IS missing, AND the first letter of the ELIG field value is NOT="N".

FY10–FY15: A patient is considered a Veteran if the following is true:
1. PRIO1_8 value is NOT missing.

(Note that in FY10, the "ELIG" field became obsolete as a Veteran identifier.)

3.2. VHA Users

VHA users were identified from ADUSH Enrollment Files using a year-specific user field labeled "FYyy", and the following cost fields:

FY00-FY07: CNHCOST; DSSCNHCOST; DSSFEECOST; DSSLTCCOST; DSSMEDCOST; DSSNVACOST; DSSOPCCOST; DSSPSYCOST; DSSSURCOST; FEECOST; LTCCOST; MEDCOST; NVACOST; OPCCOST; PSYCOST; SURCOST

FY08–FY15: ARCCNHCOST; ARCFEECOST; ARCLTCCOST; ARCMEDCOST; ARCNVACOST; ARCOPCCOST; ARCPSYCOST; ARCSURCOST; CNHCOST; FEECOST; LTCCOST; MEDCOST; NVACOST; OPCCOST; PSYCOST; SURCOST

For FY00-FY10, a person is considered to be a VHA user in a particular fiscal year if the following are true:
1. "FYyy=1" for the specified year[10], AND
2. Sum of all cost fields is >0 for the specified year.

[9] Data sources are described in Section 1.0 of this Technical Appendix.

[10] This designation in ADUSH Enrollment Files indicates that the patient appeared in a FYyy utilization file for VHA outpatient services, VHA inpatient services, VA pharmacy services, Purchased Care outpatient or inpatient services, or non-VA contract care.

For FY11-FY15, we followed the approach for the earlier cohort years and applied an additional step to exclude those Veteran patients who received VHA care exclusively for Compensation and Pension examination and/or Employee Health services.[11]

All others were non-users. The term "user" is synonymous with the term "patient" in this Sourcebook.

3.3. Sex

The sex variable algorithm used in Sourcebook Volume 4 follows the same algorithm approach as reported in Sourcebook 3. In Volume 3, WHEI revised the sex algorithm to include the most recent sex value across the source files based on the assumption that more recent values typically reflect "corrected" values.

The sex variable incorporates sex data from the ADUSH Enrollment Files, SF files, and Inpatient Main files. Creating the cross-year sex variables reported in Volume 4 involved a multi-step process.

In Step 1, we assigned the patient's sex, SEX_FINAL, based on the SEX_BEST value in the current year (FY15) ADUSH Enrollment File.

In Step 2, individuals without a SEX_FINAL value after applying Step 1 were assigned the most recent non-missing sex value from the current year (FY15) SF file.

In Step 3, individuals without a SEX_FINAL value after applying Step 2 were assigned the most recent non-missing sex value from the current year (FY15) Medical SAS Inpatient Main files.

In Step 4, for individuals without a SEX_FINAL value after applying Step 3, we repeated Steps 1-3 for FY14, and then continued to fill in missing data iteratively using the same approach by searching prior years' files in reverse year order, back to FY00.

FY06–FY15: Identified using ADUSH Enrollment File field labeled "SEX_BEST."[12]
FY00–FY05: Identified using ADUSH Enrollment File field labeled "SEX."

Together, steps 1-4 minimized missing sex values, while relying on the most recent sex data available in the ADUSH Enrollment Files and the VHA Medical SAS Datasets for FY00-FY15.

NOTE: This Sourcebook uses the term "sex" rather than "gender" because we are unable to distinguish between birth sex and self-identified gender identity (SIGI) in the data. VHA has historically collected "sex" data, which could have been interpreted as the sex recorded on a birth certificate or as the patient's self-identified gender. VHA has updated the demographic fields and now collects data on both birth sex and SIGI.

For each of the cohort years (FY00, FY05, FY10, and FY15) examined in this Sourcebook, we used the multi-year FY00-FY15 sex variable, rather than the year-specific sex variable, because more recent sex

[11] In FY15, a total of 144,814 individuals were excluded from the WHEI Master Database because their only source of VHA utilization was a visit for a Compensation and Pension disability examination or for Occupational Health services related to their VHA employment.
[12] Since FY06, the VA Information Resource Center (VIReC) Vital Status Files include derived sociodemographic fields, including SEX_BEST and DOB_BEST, which incorporate information from multiple data sources and thus represent more complete/accurate data. ADUSH Enrollment Files use these fields from FY06 onward.

values may reflect "more correct" values.[13] Among the FY15 cohort, after assigning a year-specific sex for each of the 16 years from FY00-FY15 (using the within-year hierarchy of ADUSH sex selected first, then SF sex selected only if ADUSH sex was missing, and then Medical SAS Inpatient Main sex selected only if ADUSH and SF sex were missing), only 0.2% of patients had at least one non-missing year-specific sex value different from the FY15 year-specific value.

Although the proportion of Veteran VHA users in the database with inconsistent sex values is small, accurately ascertaining whether patients are women or men is critical to the purpose of this Sourcebook. In addition to inconsistent sex values, there are a number of patients in the Base Cohorts missing sex values in the ADUSH, SF and Medical SAS Inpatient Main files. Exhibit B shows the number and proportion of patients the Base Cohort for each of the fiscal years with missing sex values.

Exhibit B. Number and Proportion of Individuals in Base Cohort Missing a Sex Value in Each Fiscal Year

Fiscal Year	Base Cohort: Total Number (Veterans only, VHA users only, women and men)	Base Cohort Members Missing a Sex Value	
		#	%
2000	3,395,205	9,082	0.3
2005	4,802,582	724	0.0
2010	5,351,873	293	0.0
2015	5,891,311	1,237	0.0

3.4. Age

Similar to the sex variable, the age variable algorithm includes the most recent date of birth (DOB) value across the source files based on the assumption that more recent values typically reflect "corrected" values. Also, the age variable in this Sourcebook incorporates date of birth data from the ADUSH Enrollment Files, SF, and Inpatient Main files. Creating the Sourcebook Volume 4 age variables involved five steps.

In Step 1, we assigned a DOB value, DOB_FINAL, based on the DOB_BEST value in the current year (FY15) ADUSH Enrollment File.

In Step 2, individuals without a DOB_FINAL value after applying Step 1 were assigned the most recent non-missing, within-range value of "DOB" from the current year (FY15) SF file.

In Step 3, individuals without a DOB_FINAL value after applying Step 2 were assigned the most recent non-missing, within-range "BORNDAY" value from the current year (FY15) Inpatient Main files.

In Step 4, for individuals without a DOB_FINAL value after applying Step 3, we repeated Steps 1-3 for FY14, and then continued to fill in missing data iteratively using the same approach by searching prior years' files in reverse year order, back to FY00.

FY06–FY15: Identified using ADUSH Enrollment Files field labeled "DOB_BEST."
FY00–FY05: Identified using ADUSH Enrollment File field labeled "DOB."

[13] Note that an unknown (but likely small) number of transgender patients may have arranged to have their sex value changed sometime between FY00 and FY15, the period examined herein.

Together, Steps 1-4 minimized missing DOB values, while relying on the most recent DOB data available in ADUSH Enrollment Files and the VHA Medical SAS Datasets for FY00-FY15.

In Step 5, we calculated age in a given year by subtracting the DOB_FINAL (identified in Steps 1-4) from the first day of the fiscal year (in days) and then dividing the result by 365.25 to determine the age in years. When this calculation resulted in a decimal, the final age value was rounded down to the nearest integer. For example, an age of 47.788 was rounded down to 47.

For each of the cohort years (FY00, FY05, FY10, and FY15) examined in this Sourcebook, we used the multi-year FY00-FY15 DOB_FINAL variable, rather than the year-specific DOB variable, because more recent DOB values may reflect "more correct" values. Among the FY15 cohort, after assigning a year-specific DOB for each of the 16 years from FY00-FY15 (using the within-year hierarchy of ADUSH DOB selected first, then SF DOB selected only if ADUSH DOB was missing, and then Medical SAS Inpatient Main DOB selected only if ADUSH and SF DOB were missing), only 1.8% of patients had at least one non-missing year-specific DOB value different from the FY15 year-specific value.

Exhibit C shows the number and proportion of patients in the Base Cohort for each of the fiscal years with missing date of birth values.

Exhibit C. Number and Proportion of Individuals in Base Cohort Missing a Date of Birth Value in Each Fiscal Year

Fiscal Year	Base Cohort: Total Number (Veterans only, VHA users only, women and men)	Base Cohort Members Missing a DOB Value	
		#	%
2000	3,395,205	9,309	0.3
2005	4,802,582	790	0.0
2010	5,351,873	385	0.0
2015	5,891,311	1,491	0.0

3.5. Race/Ethnicity Status[14]

3.5.1. Overview

Several different VHA files contain information about race/ethnicity: the VHA Medical SAS Datasets, the VA OEF/OIF/OND Roster, and the VHA VSF. By combining data across files and across years, it is possible to reduce the number of patients with missing race/ethnicity values.[15] However, race/ethnicity data is structured quite differently across data sources, or even across years within a single data source. Therefore, to make it possible to combine data from different sources, it is necessary to perform within-source, within-year data processing to achieve a standardized data structure across sources and years. We first explain how we mapped race data from different sources to a common set of response options, to be applied to our "WHEI_RACE" variable, and how we mapped ethnicity data from different sources to a common set of response options, to be applied to our "WHEI_ETHNICITY" variable. We then

[14] This description of the race/ethnicity algorithm is also described in the National Veteran Health Equity Report–FY2013. See VA Office of Health Equity. 2016. National Veteran Health Equity Report–FY2013. US Department of Veterans Affairs, Washington, DC. Available online at http://www.va.gov/healthequity/NVHER.asp. The description has been updated to reflect current data sources used in the algorithm.

[15] The VA Information Resource Center provides guidance on working with race and ethnicity data in VHA data (VA Information Resource Center. VIReC Technical Report: VA Race Data Quality. Hines, IL: U.S. Dept. of Veterans Affairs, Health Services Research and Development Service, VA Information Resource Center, Sept. 2011).

describe how we combined data within each data source and across data sources to reduce missing data as we populated the WHEI_RACE variable and the WHEI_ETHNICITY variable. Finally, we explain how we linked our WHEI_RACE variable and our WHEI_ETHNICITY variable to create a composite variable called WHEI_RACE/ETHNICITY.

Creating the race/ethnicity variable involved four phases, described next.

3.5.2. Phase 1: Mapping algorithm to standardize race and ethnicity categories across sources and across years

Since race and ethnicity classification schema are not uniform across data sources, we constructed standardized categories and mapped values from each source to these standardized categories, as detailed in Exhibits D and E. This mapping algorithm allowed us to assign standardized values to WHEI_RACE (six race categories: American Indian/Alaska Native; Asian; Black/African American; Native Hawaiian/Other Pacific Islander; White; and Unknown) and to WHEI_ETHNICITY (three ethnicity categories: Hispanic; non-Hispanic; and Unknown).

Exhibit D. Mapping of "Race" Values

Data Sources	VHA Medical SAS Datasets		VHA Medical SAS Datasets		VA OEF/OIF/OND Roster Source: Department of Defense	VHA Vital Status File Source: Medicare		WHEI_RACE Values*
Fields	RACE1-RACE7 (SF, FY04-FY15)	RACE1-RACE6 (Inpatient Main, FY03-FY15)	RACE (SF, FY00-FY03, FY04-FY15†)	RACE (Inpatient Main, FY00-FY02, FY03-FY15‡)	RACE, ETHNICITY	CMS_RACE		
Race Values from Original Source File	American Indian or Alaska Native		American Indian		**Race**=(Other OR Unknown§) AND **Ethnicity**=(Aleut, Eskimo, OR U.S./Canadian Indian tribes)	North American Native	Reassigned Race Values	**American Indian/ Alaska Native**
	Asian		Asian		**Race**=(Other OR Unknown) AND **Ethnicity**=(Asian Indian, Chinese, Filipino, Guamanian, Japanese, Korean, Vietnamese, OR Other Asian Descent)	Asian		**Asian**
	Black or African American		Hispanic Black; Black		Black	Black		**Black/ African American**
	Native Hawaiian or Other Pacific Islander				**Race**=(Other OR Unknown) AND **Ethnicity**=(Melanesian, Micronesian, Polynesian, OR Other Pacific Islander Descent)			**Native Hawaiian/ Other Pacific Islander**
	White		Hispanic White; White		White	White		**White**
	Unknown; Declined to Answer; Missing		Unknown; Missing		**Race**=Hispanic OR **Race**=(Other OR Unknown) AND **Ethnicity**=(Other, None, OR Unknown)	Hispanic; Other; Unknown		**Unknown**‖

* Within each data source within each fiscal year, during data processing we replaced the race value found in the original data source file with this reassigned WHEI_RACE value.

† Starting in FY04, RACE values in the SF file were no longer being populated, although the previously-populated legacy value was carried forward in subsequent years' files.

‡ Starting in FY03, RACE values in the Medical SAS Inpatient Main files were no longer being populated, although the previously-populated legacy value was carried forward in subsequent years' files.

§ For 0.05% of individuals in the FY15 WHEI Master Database, OEF/OIF/OND Roster Ethnicity=[(Aleut, Eskimo, OR U.S./Canadian Indian tribes) OR (Asian Indian, Chinese, Filipino, Guamanian, Japanese, Korean, Vietnamese, OR Other Asian Descent) OR (Melanesian, Micronesian, Polynesian, OR Other Pacific Islander Descent)] AND OEF/OIF/OND Roster Race=[(White) OR (Black)]. Note that these individuals' WHEI_RACE would be White or Black/African American, respectively (and not American Indian/Alaska Native, Asian or Native Hawaiian/Other Pacific Islander, respectively).

‖ Unknown includes (a) "Unknown," "Declined to Answer," or "Missing," or (b) Race coded as Hispanic without any modifier (i.e., not specified as Hispanic White or Hispanic Black), or (c) "Other" or "Unknown" Race combined with "Other," "None," or "Unknown" Ethnicity.

Exhibit E. Mapping of "Ethnicity" Values

Data Sources	VHA Medical SAS Datasets		VHA Medical SAS Datasets		VA OEF/OIF/OND Roster Source: Department of Defense	VHA Vital Status File Source: Medicare		WHEI_ ETHNICITY Values*
Fields	ETHNIC (SF, FY04-FY15)	ETHNIC (Inpatient Main, FY03-FY15)	RACE (SF, FY00-FY03, FY04-FY15†)	RACE (Inpatient Main, FY00-FY02, FY03-FY15‡)	ETHNICITY	CMS_RACE		
Ethnicity Values from Original Source File	Hispanic or Latino		Hispanic White, Hispanic Black		Puerto Rican; Mexican; Cuban; Latin American with Hispanic Descent; Other Hispanic Descent	Hispanic	Reassigned Ethnicity Values	**Hispanic**
	Not Hispanic or Latino		American Indian§		Asian Indian; Chinese; Filipino; Guamanian; Japanese; Korean; Vietnamese; Other Asian descent; Aleut; Eskimo; U.S./Canadian Indian tribes; Melanesian; Micronesian; Polynesian; Other Pacific Islander descent	North American Native§		**Non-Hispanic**
			Asian§			Asian§		
			Black			Black§		
			White			White§		
	Unknown; Declined to Answer; Missing		Unknown; Missing		Other; None; Unknown	Other; Unknown		**Unknown**‖

* Within each data source within each fiscal year, we replaced the ethnicity value found in the original data source file with this WHEI reassigned ethnicity value during data processing.

† Starting in FY04, RACE values in the SF file were no longer being populated, although the previously-populated legacy value was carried forward in subsequent years' files.

‡ Starting in FY03, RACE values in the Inpatient Main files were no longer being populated, although the previously-populated legacy value was carried forward in subsequent years' files.

§ Although it is possible that individuals with these race values could be Hispanic, WHEI mapped these to "non-Hispanic" due to the fact that "Hispanic" was a response option in these files but was not selected for the individual.

‖ Includes "Unknown," "Declined to Answer," "Missing," "Other," and "None."

3.5.3. Phase 2: Addressing missing data for race

Applying the mapping algorithm described in Exhibit D, we created a person-level race variable, WHEI_RACE, that minimized missing values by incorporating data from multiple sources and multiple years.

In Step 1, we populated WHEI_RACE with the most recent, non-missing RACE1 value in the FY15 SF file. If RACE1 was missing, we sequentially used any non-missing RACE2-RACE7 value.[16]

In Step 2, individuals without a WHEI_RACE value after applying Step 1 were assigned the most recent,

[16] The values for the RACE1-RACE7, RACE1-RACE6, and ETHNIC variables additionally contain information on the method by which race/ethnicity information was collected, i.e., whether race/ethnicity was self-identified by the patient, identified by an observer (such as a clinic clerk), identified by a proxy, or whether the method of collection of data was unknown by the patient, or missing. The WHEI race and ethnicity algorithms did not attempt to distinguish between these different data collection methods, and simply assigned a value based on the most recent, non-missing race and ethnicity values. The values "Declined to Answer" and "Unknown" were considered to be missing values.

non-missing RACE1 value in any of the FY15 Inpatient Main files. If RACE1 was missing, we sequentially used any non-missing RACE2-RACE6 value from those files.

In Step 3, individuals without a WHEI_RACE value after applying Step 2 were assigned the most recent, non-missing RACE value in the FY15 SF file.[17]

In Step 4, individuals without a WHEI_RACE value after applying Step 3 were assigned the most recent, non-missing RACE value in any of the FY15 Inpatient Main files.[18]

In Step 5, individuals without a WHEI_RACE value after applying Step 4 were assigned the RACE value from the JAN2015[19] VA OEF/OIF/OND Roster (a cumulative file), if that value was not missing or unknown.

In Step 6, individuals without a WHEI_RACE value after applying Step 5 were assigned the CMS_RACE value from the VHA VSF (a cumulative file updated in the third quarter of FY16) if that value was non-missing.

In Step 7, for individuals still without a WHEI_RACE value after applying Step 6, we repeated Steps 1-4 for FY14, and then continued to fill in missing data iteratively using the same approach by searching prior years' files in reverse year order, back to FY04. (Steps 5 and 6 did not apply to these prior years because the VA OEF/OIF/OND Roster and the VSF are cumulative files.)

In Step 8, for individuals still without a WHEI_RACE value after applying Step 7, we repeated Steps 3 and 4 for FY03, and then continued to fill in missing data iteratively using the same approach by searching prior years' files in reverse year order, back to FY00. (Steps 1 and 2 did not apply because RACE1-RACE7 were not available in FY03 and earlier SF files, and RACE1-RACE6 were not available in FY02 and earlier Inpatient Main files.)

Note that only a single WHEI_RACE value was assigned to each patient. However, for patients whose WHEI_RACE value was assigned based on the RACE1-RACE7 or RACE1-RACE6 fields in the VHA Medical SAS Datasets, a person-level variable was created to count the number of valid race values that appeared across all race fields in the same record of the patient's most recent, non-missing Race1-RACE7 value in the SF file or in the same record of the patient's most recent, non-missing Race1-RACE6 value in the Inpatient Main files.

3.5.4. Phase 3: Addressing missing data for ethnicity

Applying the mapping algorithm described in Exhibit E, we likewise created a person-level ethnicity variable, WHEI_ETHNICITY, that minimized missing values by incorporating data from multiple sources and multiple years.

In Step 1, we populated WHEI_ETHNICITY with the most recent, non-missing ETHNIC value in the FY15 SF file.

[17] Note that starting in FY04, RACE values in the SF file were no longer being populated, although the previously-populated legacy value was carried forward in subsequent years' files.

[18] Note that starting in FY03, RACE values in the Inpatient Main files were no longer being populated, although the previously-populated legacy value was carried forward in subsequent years' files.

[19] Note that this is a partial fiscal year file. As of the time analyses were completed in 2017, more recent Roster data were not available to WHEI. Although only a small proportion of patients (1%) are assigned a race/ethnicity value at this step, assignment of race/ethnicity at this step may disproportionately affect Veterans who have recently returned from military service.

In Step 2, individuals without a WHEI_ETHNICITY value after applying Step 1 were assigned the most recent, non-missing ETHNIC value in any of the FY15 Inpatient Main files.

In Step 3, individuals without a WHEI_ETHNICITY value after applying Step 2 were assigned an ethnicity value from the most recent, non-missing RACE value in the FY15 SF files.[20]

In Step 4, individuals without a WHEI_ETHNICITY value after applying Step 3 were assigned an ethnicity value from the most recent, non-missing RACE value in any of the FY15 Inpatient Main files.[21]

In Step 5, individuals without a WHEI_ETHNICITY value after applying Step 4 were assigned an ethnicity value from the ETHNICITY value from the JAN2015 VA OEF/OIF/OND Roster (a cumulative file) if that value was not missing or unknown.

In Step 6, individuals without a WHEI_ETHNICITY value after applying Step 5 were assigned the ethnicity value from the CMS_RACE field from the FY16Q3 VHA VSF (a cumulative file), if that value was non-missing.

In Step 7, for individuals without a WHEI_ETHNICITY value after applying Step 6, we repeated Steps 1-4 for FY14, and then continued to fill in missing data iteratively using the same approach by searching prior years' files in reverse year order, back to FY04. (Steps 5 and 6 did not apply to these prior years because the VA OEF/OIF/OND Roster and the VSF are cumulative files.)

In Step 8, for individuals without a WHEI_ETHNICITY after applying Step 7, we repeated steps 3 and 4 for FY03, and then continued to fill in missing data iteratively using the same approach by searching prior years' files in reverse year order, back to FY00. (Steps 1 and 2 did not apply because ETHNIC was not available in FY03 and earlier SF files nor in FY02 and earlier Inpatient Main files.)

Note that only a single WHEI_ETHNICITY value was assigned to each patient.

3.5.5. Phase 4: Creating a combined, person-level race/ethnicity variable

Finally, we combined our person-level WHEI_RACE variable with our person-level WHEI_ETHNICITY variable to create a single, person-level WHEI_RACE/ETHNICITY variable, using the mapping strategy described in Exhibit F. This mapping is adapted from the approach used by The Statistical Policy Division, Office of Information and Regulatory Affairs, of the Office of Management and Budget, which parallels one approach used for U.S. Census data.[22,23] The approach generally matches the approach used in prior WHEI reports but includes a new "Multi-race" category to be consistent with the VHA Office of Health Equity classification approach.[24]

[20] RACE contains legacy data from prior to 2004 when race and ethnicity were reported in the same variable. Note that starting in FY04, RACE values in the SF file were no longer being populated, although the previously-populated value was carried forward in subsequent years' files.
[21] Note that starting in FY03, RACE values in the Inpatient Main files were no longer being populated, although the previously-populated value was carried forward in subsequent years' files.
[22] Office of Management and Budget, Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity, 30 October 1997, https://www.gpo.gov/fdsys/pkg/FR-1997-10-30/pdf/97-28653.pdf (Accessed Jan 2, 2018).
[23] 2010 Census Redistricting Data (Public Law 94-171) Summary File—Technical Documentation/prepared by the U.S. Census Bureau, 2011.
[24] VA Office of Health Equity. 2016. National Veteran Health Equity Report–FY2013. US Department of Veterans Affairs, Washington, DC. Available online at http://www.va.gov/healthequity/NVHER.asp.

Exhibit F. Mapping of WHEI_RACE and WHEI_ETHNICITY to WHEI_RACE/ETHNICITY

WHEI_RACE (From Exhibit D)	WHEI_ETHNICITY (From Exhibit E)	WHEI_RACE/ETHNICITY (combined)
American Indian/Alaska Native	Hispanic	**Hispanic**
Asian	Hispanic	
Black/African American	Hispanic	
Native Hawaiian/Other Pacific Islander	Hispanic	
White	Hispanic	
Unknown	Hispanic	
Any combination of 2 or more races recorded in the same record	Non-Hispanic OR Unknown	**Multi-race**
American Indian/Alaska Native	Non-Hispanic OR Unknown	**American Indian/Alaska Native – non-Hispanic**
Asian	Non-Hispanic OR Unknown	**Asian – non-Hispanic**
Black/African American	Non-Hispanic OR Unknown	**Black/African American – non-Hispanic**
Native Hawaiian/Other Pacific Islander	Non-Hispanic OR Unknown	**Native Hawaiian/Other Pacific Islander**
White	Non-Hispanic OR Unknown	**White – non-Hispanic**
Unknown	Non-Hispanic OR Unknown	**Unknown**

Note: All individuals with indication of Hispanic ethnicity are included in the "Hispanic" race/ethnicity group regardless of their race. The remaining race/ethnicity categories contain Veteran patients who have identified as "non-Hispanic," but for simplicity, the labels reported in the main text identifies only the race. For example, in the text, "American Indian/Alaska Native – non-Hispanic" is shortened to "American Indian/Alaska Native."

We used a hierarchical approach to assign the final person-level combined WHEI_RACE/ETHNICITY values.

In Step 1, among individuals whose WHEI_ETHNICITY value was Hispanic, we automatically assigned a Hispanic WHEI_RACE/ETHNICITY value, regardless of the WHEI_RACE value. For example, if an individual had a WHEI_RACE value of "Black/African American" but his/her WHEI_ETHNICITY value was "Hispanic", his/her WHEI_RACE/ETHNICITY value was assigned as "Hispanic."

In Step 2, among the individuals remaining, we then looked at records for the subset of individuals whose WHEI_RACE was assigned based on the RACE1-RACE7 fields in the SF file.

- In Step 2a, if an individual had two or more values of RACE recorded in the same record as his/her most recent, non-missing RACE1-RACE7 value, then the individual was assigned a "MULTI-RACE" value for WHEI_RACE/ETHNICITY.
- In Step 2b, among the remaining individuals whose WHEI_RACE was assigned based on the RACE1-RACE7 fields in the SF file, we assigned the same WHEI_RACE/ETHNICITY value as the WHEI_RACE value. For example, an individual with a WHEI_RACE value of "White" was then assigned a WHEI_RACE/ETHNICITY value of "White."

In Step 3, among the individuals missing WHEI_RACE/ETHNICITY values at the end of Step 2, we next looked at records for the subset of individuals whose WHEI_RACE was assigned based on the RACE1-RACE6 fields in the Inpatient Main files.

- In Step 3a, if an individual had two or more values of RACE recorded in the same record as his/her most recent, non-missing RACE1-RACE6 value, then the individual was assigned a "MULTI-RACE" value for WHEI_RACE/ETHNICITY.
- In Step 3b, among the remaining individuals whose WHEI_RACE was assigned based on the RACE1-RACE6 fields in the Inpatient Main files, we assigned the same WHEI_RACE/ETHNICITY value as the WHEI_RACE value. For example, an individual with a WHEI_RACE value of "American Indian/Alaska Native" was then assigned a WHEI_RACE/ETHNICITY value of "American Indian/Alaska Native."

In Step 4, of the remaining individuals with missing WHEI_RACE/ETHNICITY values at the end of Step 3, individuals with "Unknown" WHEI_RACE were mapped to "Unknown" WHEI_RACE/ETHNICITY. All others with known WHEI_RACE were mapped to the corresponding non-Hispanic category, as described in Exhibit F. Exhibit G shows the number and proportion of patients in each fiscal year who were missing race/ethnicity values, by sex.

Exhibit G. Number and Proportion of Individuals in Base Cohort Missing a WHEI Race/Ethnicity Value in Each Fiscal Year, by Sex

Fiscal Year	Women			Men		
	N	# Missing	% Missing	N	# Missing	% Missing
2000	159,810	12,965	8.1	3,226,313	96,210	3.0
2005	231,907	12,170	5.3	4,569,951	87,947	1.9
2010	317,122	13,368	4.2	5,034,458	92,275	1.8
2015	439,791	21,895	5.0	5,450,283	152,709	2.8

For each of the cohort years (FY00, FY05, FY10, and FY15) examined in this Sourcebook, we used the FY15 race/ethnicity variable, rather than the cohort year-specific race/ethnicity variable, because more recent race/ethnicity values may reflect "more correct" values.

3.6. Urban/Rural Status

3.6.1. Overview

The urban/rural variable algorithm used in Sourcebook Volume 4 differs from the algorithms described in previous volumes because of data availability. The urban/rural variable reported in Volumes 2 and 3 drew from the "URH" field in the Planning Systems Support Group (PSSG) Enrollee Files, which indicate the urban/rural status of the last known address in the fiscal year for each enrollee, and was supplemented with population data from the US Census. Because the PSSG file is only available starting in FY09, for Sourcebook Volume 4 we applied a different approach, using zip codes, to characterize urban/rural status for the 16-year time interval reflected in this volume.

Creating the urban/rural variable involved three phases.

3.6.2. Phase 1: Assignment of zip code values across multiple data sources

For most sociodemographic variables, we start first with the ADUSH Enrollment file. However, we used the utilization files as the data source for patients' residential zip codes. We searched for zip code values across multiple data sources in order to reduce missingness.

For each patient in the WHEI cohort, we assigned zip code values from the most-recent, non-missing record in the utilization files using the hierarchy below, with an indicator for zip code source.

1. SF
2. Inpatient Main Files
3. Purchased Care Outpatient Services File
4. Purchased Care Inpatient File
5. Purchased Care Inpatient Ancillary File

3.6.3. Phase 2: Approximating zip codes to Rural/Urban Commuting Area (RUCA) codes

The RUCA codes are a U.S. Census-based classification scheme that uses measures of population density and daily commuting patterns to characterize rural and urban areas. The WWAMI Rural Health Research Center at the University of Washington[25] has developed zip code RUCA approximations using Census commuting data and zip codes.[26] To date, they have created three versions:

1. Version 1.11 uses 1998 Zip codes and 1990 Census commuting data
2. Version 2.0 has a version that uses 2004 Zip codes and 2000 Census commuting data and an updated version that uses 2006 Zip codes and 2000 Census commuting data
3. Version 3.10 uses 2013 Zip codes and 2010 Census commuting data

To create the urban/rural variable for each fiscal year reported, we used the following versions:

FY00: Version 2.0 (2006 Zip code version)
FY05: Version 2.0 (2006 Zip code version)
FY10: Version 3.10 (2013 Zip code version)
FY15: Version 3.10 (2013 Zip code version)

3.6.4. Phase 3: Defining the mapping strategy of RUCA codes to urban/rural categories

The mapping system of RUCA codes to urban/rural categories used in this report to describe population density follows the definitions used by PSSG and the Office of Rural Health (ORH).[27] Exhibit H describes the mapping of RUCA codes to the urban/rural categories of urban, rural, and highly rural.

[25] See http://depts.washington.edu/uwruca/index.php.
[26] RUCA codes based on U.S. census tracts are also available. For more information, see https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx.
[27] See https://www.ruralhealth.va.gov/aboutus/ruralvets.asp.

Exhibit H. Mapping of RUCA Codes to Urban/Rural Categories

RUCA V2.0	RUCA V3.10	RUCA Code Definitions (Zip Code version)	WHEI Urban/Rural Category
		1 Metropolitan area core: primary flow within an Urbanized Area	
x	x	1.0 No additional code	Urban
x	x	1.1 Secondary flow 30% through 49% to a larger Urbanized Area	Urban
		2 Metropolitan area high commuting: primary flow 30% or more to an Urbanized Area	
x	x	2.0 No additional code	Rural
x	x	2.1 Secondary flow 30% through 49% to a larger Urbanized Area	Rural
		3 Metropolitan area low commuting: primary flow 10% to 30% to an Urbanized Area	
x	x	3.0 No additional code	Rural
		4 Micropolitan area core: primary flow within an Urban Cluster of 10,000-49,999 (large Urban Cluster)	
x	x	4.0 No additional code	Rural
x	x	4.1 Secondary flow 30% through 49% to an Urbanized Area	Rural
x		4.2 Secondary flow 10% through 29% to an Urbanized Area	Rural
		5 Micropolitan high commuting: primary flow 30% or more to a large Urban Cluster	
x	x	5.0 No additional code	Rural
x	x	5.1 Secondary flow 30% through 49% to an Urbanized Area	Rural
x		5.2 Secondary flow 10% through 29% to an Urbanized Area	Rural
		6 Micropolitan low commuting: primary flow 10% to 30% to a large Urban Cluster	
x	x	6.0 No additional code	Rural
x		6.1 Secondary flow 10% through 29% to an Urbanized Area	Rural
		7 Small town core: primary flow within an Urban Cluster of 2,500-9,999 (small Urban Cluster)	
x	x	7.0 No additional code	Rural
x	x	7.1 Secondary flow 30% through 49% to an Urbanized Area	Rural
x	x	7.2 Secondary flow 30% through 49% to a large Urban Cluster	Rural
x		7.3 Secondary flow 10% through 29% to an Urbanized Area	Rural
x		7.4 Secondary flow 10% through 29% to a large Urban Cluster	Rural
		8 Small town high commuting: primary flow 30% or more to a small Urban Cluster	
x	x	8.0 No additional code	Rural
x	x	8.1 Secondary flow 30% through 49% to an Urbanized Area	Rural
x	x	8.2 Secondary flow 30% through 49% to a large Urban Cluster	Rural
x		8.3 Secondary flow 10% through 29% to an Urbanized Area	Rural
x		8.4 Secondary flow 10% through 29% to a large Urban Cluster	Rural
		9 Small town low commuting: primary flow 10% through 29% to a small Urban Cluster	
x	x	9.0 No additional code	Rural
x		9.1 Secondary flow 10% through 29% to an Urbanized Area	Rural
x		9.2 Secondary flow 10% through 29% to a large Urban Cluster	Rural
		10 Rural areas: primary flow to a tract outside an Urbanized Area or Urban Cluster (including self)	
x	x	10.0 No additional code	Highly Rural

RUCA		RUCA Code Definitions (Zip Code version)	WHEI Urban/Rural Category
V2.0	V3.10		
x	x	10.1 Secondary flow 30% through 49% to an Urbanized Area	Rural
x	x	10.2 Secondary flow 30% through 49% to a large Urban Cluster	Rural
x	x	10.3 Secondary flow 30% through 49% to a small Urban Cluster	Rural
x		10.4 Secondary flow 10% through 29% to an Urbanized Area	Rural
x		10.5 Secondary flow 10% through 29% to a large Urban Cluster	Rural
x		10.6 Secondary flow 10% through 29% to a small Urban Cluster	Rural

Notes: An "x" in the RUCA V2.0 and V3.10 columns indicate that the RUCA code is included in the corresponding version. A black box indicates that the RUCA code was not available in the corresponding version.
Source: RUCA Data Code Definition: Version 2.0. Available from: http://depts.washington.edu/uwruca/ruca-codes.php

For a geographic representation of the assignment of urban/rural categories to zip codes in the contiguous United States, please see the Online Appendix (Supplemental Materials).

Exhibit I shows the number and proportion of patients who were missing urban/rural values in each fiscal year.

Exhibit I. Number and Proportion of Individuals in Base Cohort Missing an Urban/Rural Value in Each Fiscal Year

Fiscal Year	Base Cohort: Total Number (Veterans only, VHA users only, women and men)	Base Cohort Members Missing an Urban/Rural Value	
		#	%
2000	3,395,205	137,549	4.1
2005	4,802,582	118,207	2.5
2010	5,351,873	134,582	2.5
2015	5,891,311	148,513	2.5

3.7. Service-Connected (SC) Disability Rating

The SC disability rating variable is based on the field "SCPER" in the ADUSH Enrollment File. Like the Veteran field, SCPER can potentially change across years for legitimate reasons (i.e., if the individual's SC disability rating changes). If the SCPER field was populated in the ADUSH Enrollment File for a particular fiscal year, we assigned the ADUSH Enrollment File SCPER value to the individual for that fiscal year. If the SCPER field was missing for that fiscal year, we considered the individual as not having an SC disability rating in that fiscal year. We created a variable indicating whether the individual had an SC disability rating in the fiscal year being examined (yes/no). For those who did have an SC disability rating in the fiscal year, we also created a variable indicating the level of the SC disability rating in that fiscal year: 0-49 percent disability rating, 50-99 percent disability rating, or 100 percent disability rating.[28]

Exhibit J shows the number and proportion of patients who were missing SC values in each fiscal year.

[28] Note that "0 percent" refers to a patient who does have an SC disability rating, but whose severity rating is 0 (zero) percent; this is distinct from a patient who has no SC disability rating.

Exhibit J. Number and Proportion of Individuals in Base Cohort Missing Service-Connected Disability Ratings in Each Fiscal Year

Fiscal Year	Base Cohort: Total Number (Veterans only, VHA users only, women and men)	Base Cohort Members Missing an SC Value	
		#	%
2000	3,395,205	34,054	1.0
2005	4,802,582	59,064	1.2
2010	5,351,873	8,058	0.2
2015	5,891,311	2,963	0.1

4.0 Algorithms for Part 2: Outpatient Utilization

4.1. Overview

We first describe the steps taken to create the VHA outpatient utilization variables, including specifications for the types of VHA outpatient care reported in this Sourcebook, and then describe the process of creating the Purchased care outpatient utilization variable.

4.2. VHA Outpatient Utilization

Outpatient utilization variables are derived from the Medical SAS VHA Outpatient Event (SE) files.

4.2.1. Generating count of VHA outpatient encounters

The WHEI Master Database contains variables counting the number of VHA *encounters* a patient had within a specific type of care. Clinic "stop codes" (codes indicating clinic type) identify the clinical setting in which the patient received care.[29] This report examines the following specific types of outpatient care:

a. **Total Outpatient Care** refers to any type of outpatient care (i.e., all clinic stop codes are considered outpatient care).[30]
b. **Total Primary Care** refers to primary care received in general medical clinics or in Women's Health Clinics.
    i. **General Primary Care Clinic** refers to primary care received in a general medical clinic or equivalent setting. In addition to providing preventive care and care for gender-neutral conditions, such clinics sometimes provide gender-specific care to women (such as cervical cancer screening and breast exams), and sometimes refer women to a different clinic for gender-specific services.
    ii. **Women's Health Clinic** refers to primary care services received in a clinic designed specifically for women. Such clinics provide comprehensive primary care services to women (i.e., preventive health care, care for gender-neutral conditions, and care for gender-specific conditions). *Note:* Historically, prior to FY10, stop code 322[31] was sometimes used to describe care for gender-specific conditions (such as cervical cancer screening and breast exams) provided in a clinic (e.g., a "pap clinic") for women receiving most of their primary care in non-gender-specific general primary care clinics. While the VHA coding manual now defines this type of clinic as 704, some facilities may not yet have fully implemented this coding change.
c. **Mental Health/Substance Use Disorder (SUD) Care** refers to care received in mental health or SUD clinics (e.g., psychiatry visits, psychology visits, individual or group therapy, SUD treatment, and mental health/SUD rehabilitation treatment programs). It also includes visits with mental health providers embedded in primary care settings. *Note:* This category does not include

[29] "Stop codes" are clinic type codes, which are used to identify outpatient care in VHA. Each type of clinic has a unique three-digit code. The codes are entered into the local VHA VISTA system for each patient encounter (e.g., a clinic visit, a radiology procedure, a clinical telephone encounter). The data gathered through VISTA are aggregated into SE files in the national SAS Medical Datasets.

[30] Previous Sourcebook volumes describe the proportion of patients with any face-to-face VHA outpatient encounters. For the proportion of outpatients with a face-to-face encounter in the VHA outpatient setting in FY12, see Sourcebook Volume 3.

[31] Starting in FY10 and continuing through FY15 (with minor revisions), the women's health clinic stop code (322) is officially described in the FY15 VHA coding manual as follows: "Records patient visit for primary care services provided to women through a coordinated, interdisciplinary provision of medical, nursing, psychosocial, and allied health services for disease treatment and prevention and health promotion and education, referral for specialty, rehabilitation, and other levels of care, follow-up and overall care management by a Comprehensive Women's Health Primary Care Provider and support team within a separate women's clinic or Center. This includes Comprehensive Women's Center (model 3) and Separate but Shared Space Women's Clinics (model 2). Additional details on models of care are available in Handbook 1300.01 'Health Care Services for Women Veterans.' Includes provider and support services. Subspecialty services may also be provided in the same physical location."

services provided by primary care providers for mental health conditions or SUDs. Screening for these conditions occurs in primary care settings, and patients may receive pharmacotherapy or brief interventions for these conditions from primary care providers as well. Also note that the mental health/SUD category does not include services provided in Social Work Clinic.[32]

For each type of care, the WHEI Master Database contains a count of the total number of encounters occurring for a patient in one fiscal year (for each year FY00– FY15), regardless of whether those encounters occurred on the same day. Of note, while we exclude duplicate records (encounters by the same person on the same day at the same facility to the same clinic stop code), more than one encounter may legitimately occur on a single day. For example, a patient may visit a primary care clinic, cardiology clinic, podiatry clinic, and the outpatient laboratory all on the same day. It is important to capture all visits occurring on each day (rather than simply counting total number of days on which care was received), because some patients try to schedule as much care as possible on a single day (e.g., to minimize travel to the care setting or to minimize time away from work or care giving).

4.2.2. VHA outpatient care variables stop codes

The specific clinic stop codes from the SE file (CL field, unless otherwise indicated in a footnote) used to create counter variables for each type of care are listed here.

Outpatient Care Encounters

Any stop code in the CL field (CL captures primary stop codes).

Primary Care Clinics (Other Than Women's Health Clinic)[33]

Due to changes over time in the delivery of primary care, we used different primary care clinic definitions for each fiscal year (as indicated by an "x" in Exhibit K).

Exhibit K. Primary Care Clinic Stop Codes

Code	Label	FY00	FY05	FY10	FY15
160323	Clinical Pharmacy - Primary Care[34]	x	x	x	x
170	HBPC - Physician	x	x	x	x
171	HBPC - Nursing (RN or LPN)	x	x	x	x
172	HBPC - Physician Extender (NP, CNS, PA)	x	x	x	x
301	General Internal Medicine	x	x	x	x
310323	Infectious Disease Primary Care[35]	x	x	x	x
318	Geriatric Problem-Focused Clinic	x	x	x	x
323	Primary Care Medicine[36]	x	x	x	x
342	Family Practice				x
348	Primary Care Shared Appointment		x	x	x

[32] This is a change from Sourcebook Volume 2, which counted Social Work Clinic as one of the mental health/SUD stop codes. This modification was made because further investigation for Sourcebook Volume 3 revealed that the preponderance of Social Work Clinic encounters were associated with a diagnosis indicative of a need for social services (such as housing instability or employment difficulties) rather than a mental health diagnosis. Thus, services provided by a social work clinician would count as mental health/SUD care if associated with a mental health/SUD clinic stop code, but services provided by a social work clinician within a Social Work clinic would not.

[33] Stop code labels come from VA documentation, and so in general no effort has been made here to spell out these abbreviations.

[34] 160323 is a code combining two different clinics, where 160 is the primary stop code (in the CL field) and 323 is the secondary stop code (in the CLC field).

[35] 3103323 is a code combining two different clinics, where 310 is the primary stop code (in the CL field) and 323 is the secondary stop code (in the CLC field).

[36] 323 is the stop code most commonly used for primary care clinics.

Code	Label	FY00	FY05	FY10	FY15
350	GeriPACT	x	x	x	x
563	MH Primary Care Team - Group	x	x		
704	Women's Gender-Specific Preventive Care[37]	x	x	x	x

Women's Health Clinics

Any of the following stop codes in the CL field (unless otherwise indicated in a footnote):

160322 Clinical Pharmacy Comprehensive Women's Primary Care Clinic[38]
322 Comprehensive Women's Primary Care Clinic

Mental Health/Substance Use Disorder Care Clinics

Due to changes over time in the delivery of mental health/SUD care, we used different definitions for each fiscal year (as indicated by an "x" in Exhibit L).

Exhibit L. Mental Health/SUD Clinic Stop Codes

Code	Label	FY00	FY05	FY10	FY15
156	HBPC - Psychologist			x	x
157	HBPC - Psychiatrist			x	x
292	Observation Psychiatry	x	x	x	x
502	Mental Health Clinic - Individual	x	x	x	x
503	Mental Health Residential Care - Individual	x	x	x	x
505	Day Treatment - Individual	x	x	x	x
506	Day Hospital - Individual	x	x	x	x
509	Psychiatry - Individual	x	x	x	x
510	Psychology - Individual	x	x	x	x
512	Mental Health Consultation	x	x	x	x
513	Substance Use Disorder Individual	x	x	x	x
514	Substance Use Disorder - Home Visit	x	x	x	x
516	PTSD - Group	x	x	x	x
519	Substance Use Disorder/PTSD Teams	x	x	x	x
523	Opioid Substitution	x	x	x	x
524	Active Duty Sexual Trauma	x	x	x	x
525	Women's Stress Disorder Treatment Teams	x	x	x	x
529	HCHV/HCMI- Individual	x	x	x	x
532	Psychosocial Rehabilitation - Individual	x	x	x	
533	Mental Health Intervention Biomedical Care- Individual		x	x	x
534	Mental Health Integrated Care - Individual			x	x
535	Mental Health Vocational Assistance Individual	x	x	x	x
539	MH Integrated Care Group			x	x
540	PTSD Clinical Team (PCT) Post-Traumatic Stress Ind	x	x	x	x
547	Intensive Substance Use Disorder - Group	x	x	x	x

[37] While this clinic provides care to women only, it is included in primary care clinics (rather than women's health clinics) because it represents only one element of primary care; in this report, "Women's Health Clinic" is reserved for care provided as part of *comprehensive* women's health care.
[38] 160322 is a code combining two different clinics, where 160 is the primary stop code (in the CL field) and 322 is the secondary stop code (in the CLC field).

Code	Label	FY00	FY05	FY10	FY15
548	Intensive Substance Use Disorder - Individual			x	x
550	Mental Health Clinic - Group	x	x	x	x
552	Mental Health Intensive Case Management (MHICM)- Individual	x	x	x	x
553	Day Treatment - Group	x	x	x	x
554	Day Hospital - Group	x	x	x	x
557	Psychiatry - Group	x	x	x	x
558	Psychology - Group	x	x	x	x
559	Psychosocial Rehabilitation - Group	x	x	x	
560	Substance Use Disorder - Group	x	x	x	x
561	PCT-Post Traumatic Stress - Group	x	x	x	x
562	PTSD - Individual	x	x	x	x
564	Mental Health Team Case Management	x	x	x	x
567	MHICM - Group		x	x	x
568	Mental Health Compensated Work Therapy/Supported Employment (CWT/SE) Face-to-Face		x	x	x
571	SeRV-MH (Services for Returning Veterans-Mental Health) - Individual			x	x
572	SeRV-MH - Group			x	x
573	Mental Health Incentive Therapy Face-to-Face	x	x	x	x
574	Mental Health Compensated Work Therapy/Transitional Work Experience (CWT/TWE) Face-to-Face	x	x	x	x
575	Mental Health Vocational Assistance - Group	x	x	x	x
576	Psycho-Geriatric Clinic - Individual	x	x	x	x
577	Psycho-Geriatric Clinic - Group	x	x	x	x
580	PTSD Day Hospital	x	x	x	x
581	PTSD Day Treatment	x	x		
582	Psychosocial Rehabilitation Recovery Center (PRRC) - Individual			x	x
583	PRRC - Group			x	x
586	Residential Rehabilitation Treatment Program (RRTP) - Individual				x
587	RRTP - Group				x
588	RRTP Aftercare - Individual			x	x
589	Non-Active Duty Sexual Trauma	x	x		
590	Community Outreach to Homeless Veterans by Staff Other than HCHV and RRTP Programs	x	x	x	x
591	Incarcerated Veterans Re-Entry			x	x
592	Veterans Justice Outreach			x	x
593	RRTP Outreach Services			x	x
594	RRTP Aftercare - Community			x	
595	RRTP Aftercare - Group			x	x
596	RRTP Admission Screening Services			x	x
598	RRTP Pre-Admission - Individual			x	x
599	RRTP Pre-Admission - Group			x	x
725	Domiciliary Outreach Services	x	x		
726	Domiciliary Aftercare - Community	x	x		
727	Domiciliary Aftercare - VA	x	x		

Code	Label	FY00	FY05	FY10	FY15
728	Domiciliary Admission Screening Services	x	x		
730	Domiciliary - General Care [Event Capture System (ECS) Use Only]	x	x		
731	Psychiatric Resident Rehabilitation Treatment Program (PRRTP) - General Care (ECS Use Only)	x	x		

4.3. Purchased Care Outpatient Utilization

4.3.1. Overview

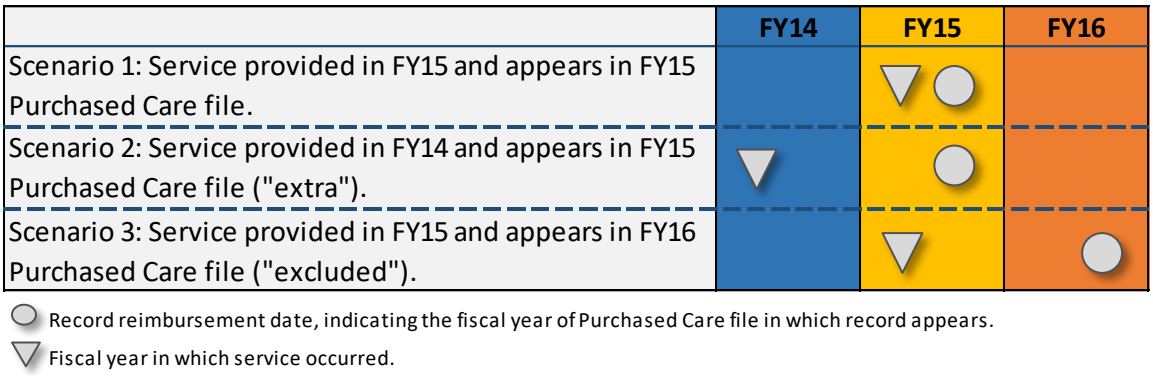
Purchased care outpatient utilization variables are derived using the Purchased Care Outpatient Services file. For each fiscal year examined, this Sourcebook reports the proportion of Veterans who had at least one outpatient service reimbursed through Purchased Care in that fiscal year.[39] In some cases, a service reimbursed in a particular year actually reflects care received in a previous year. For example, the FY15 Purchased Care Outpatient Services file includes only services that were *reimbursed* by VHA in FY15. Exhibit M shows three possible combinations of the year in which a service was provided and the year in which the service was reimbursed (and thus appeared in Purchased Care outpatient data).

Scenario 1 shows a service both provided and reimbursed in FY15.

Scenario 2 shows a service provided in FY14 but which appeared in the FY15 Purchased Care Outpatient Services file rather than FY14 Purchased Care Outpatient Services file due to a lag between service provision and service reimbursement,[40] and thus we refer to it as an "extra" service in the FY15 Purchased Care Outpatient Services file.

Scenario 3 shows a similar lag, where the service was provided in FY15 but was reimbursed in FY16; this service appears in the FY16 file but not in the FY15 file, and thus we refer to it as being "excluded" from the FY15 Purchased Care file.

Exhibit M. Three Scenarios Observed in FY15 Purchased Care Outpatient Data



[39] In prior Sourcebook Volumes, we describe steps for cleaning the Purchased Care Outpatient Services file in order to generate days and counts of Purchased Care services (see Sourcebook Volume 3 Technical Appendix, Section 9.6.2 for the data processing of Non-VA (Fee) Medical Care Outpatient Services file, available at: http://www.womenshealth.va.gov/WOMENSHEALTH/docs/Sourcebook_Vol_3_FINAL.pdf). Since this volume only describes any use of outpatient services in Purchased Care (1+ instance), the additional data processing steps were not necessary, and are thus, not described here.

[40] In the raw FY15 Purchased Care outpatient file, 77% of records reflect FY15 care, 22% reflect FY14 care, and 1% reflect care prior to FY14.

WHEI decided to create Purchased Care utilization variables based on care reimbursed in FY15, with two justifications: First, for administrative purposes, it may be useful to track the volume of services that were reimbursed in FY15, rather than the services that were provided in FY15. Second, for program evaluation purposes, volume of services *reimbursed* in FY15 appears to be an acceptable proxy for services *provided* in FY15. WHEI estimates that the number of "extra" services in the FY15 file (those provided prior to FY15 but reimbursed in FY15) will approximately compensate for the "excluded" services in the FY15 file (those provided in FY15 but reimbursed after FY15). However, annual increases in the numbers of Veterans in VHA and corresponding increases in service volume each year may mean that the number of FY15 services "excluded" exceeds the number of "extra" FY14 services included. Therefore, approximating services provided in FY15 using the FY15 Purchased Care outpatient data most likely undercounts the services actually provided in Purchased Care in FY15. This could be particularly relevant to FY15 data, because in FY15, the Purchased Care file included not only records of fee basis care, but also records of care under the Veterans Choice Program. There may have been delays in payment of some FY15 Veterans Choice Program care.

4.3.2. Generating indicator for any use of Purchased Care outpatient services

The WHEI Master Database contains a person-level variable indicating any use (1+ instance) of Purchased Care outpatient services if the patient had any encounter recorded in the FY15 Purchased Care Outpatient Services file.

5.0 Algorithms for Part 2: Reproductive Health Care

5.1. Overview

New to Sourcebook Volume 4 is the characterization of Reproductive Health care. This Sourcebook reports on two types of Reproductive Health care: VHA obstetrics/gynecology specialty care provided within the VHA and/or Purchased Care outpatient setting, and VHA-covered obstetric deliveries provided through the Purchased Care inpatient setting.

5.2. Obstetrics/Gynecology Specialty Care

We first defined obstetrics/gynecology specialty care in both the VHA and Purchased Care outpatient settings.

- **VHA outpatient:** The VHA outpatient event (SE) file contains clinic stop codes that specify the type of care a patient receives (see Section 4.2.2). VHA obstetrics/gynecology specialty care refers to the presence of at least one encounter during the year occurring in VHA gynecology clinics (clinic stop codes 404 and /or 426) and/or (starting in FY15) in VHA obstetrics clinics (clinic stop code 339).
- **Purchased Care outpatient:** Purchased Care obstetrics/gynecology specialty care refers to the presence of at least one outpatient record—during the year being examined—in the Purchased Care system, for which the reason for the visit is a diagnosis code in the "pregnancy" or other "reproductive health" category.[41] These conditions include 1+ instance of any of the following conditions recorded in the Purchased Care Outpatient Services file: Sexually Transmitted Infections, Vaginitis and Other Pelvic Inflammatory Conditions, Cervical Dysplasia/ Atypical Squamous Cells of Uncertain Significance (ASCUS), Endometriosis, Menstrual Disorders, Fibroids, Ovarian Cyst, Polycystic Ovaries, Benign Gynecologic Neoplasms – Other, Prolapse of Female Genital Organs, Reproductive Organ Disorders – Other, Sexual Dysfunction, Contraceptive Care Management, Infertility, Menopausal Disorders, Miscarriage, Ectopic Pregnancy, Pregnancy or Delivery – Normal, Pregnancy with Obstetrical Complications or Prolonged, Pregnancy Complicated by Diabetes Mellitus, Pregnancy Complicated by Hypertension, Pregnancy Complicated by Other Medical Conditions, Cancer – Cervical, Cancer – Uterine, Cancer – Ovarian, Cancer - Female Reproductive – Other, Carcinoma in Situ – Cervical, and Carcinoma in Situ - Female Reproductive – Other.[42] See Section 6.0 for the algorithm for creating medical and mental health conditions and domains.

To determine obstetrics/gynecology specialty care use by setting, we created a person-level database with no/yes (0/1) indicators for obstetrics/gynecology specialty care:

a. VHA obstetrics/gynecology specialty care
b. Purchased Care obstetrics/gynecology specialty care

Finally, we assigned women Veteran patients to one of four mutually exclusive groups using the indicators above:

1. Both VHA and Purchased Care obstetrics/gynecology specialty care (a=1 AND b=1)
2. VHA obstetrics/gynecology specialty care only (a=1 AND b =0)
3. Purchased Care obstetrics/gynecology specialty care only (a=0 AND b=1)

[41] The Purchased Care Outpatient Services files do not contain an equivalent field for clinic type; therefore, we used diagnosis of an obstetrics/gynecology-related condition recorded in the Purchased Care outpatient file as a proxy for obstetrics/gynecology specialty care.
[42] Male Genital Disorders are also part of the "reproductive health" category, but do not apply to women.

4. No obstetrics/gynecology specialty care (a=0 AND b=0)

5.3. VHA-Covered Obstetric Deliveries

To create the obstetric deliveries variable, WHEI adapted a validated algorithm[43,44] and applied updates to capture all deliveries covered by VHA over the 16-year period. The obstetric deliveries variable draws upon multiple fields in the Purchased Care Inpatient Stay file:

1. International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis codes (variables DXLSF, DX2-DX5)
2. Pricer Diagnostic Related Groups (DRG) (variable PDRG)
3. Surgery procedure codes (variables SURG9CD1-SURG9CD5)

Additionally, records were excluded if there was indication of pregnancies that were non-viable (i.e. ectopic or molar pregnancies) and/or ended in early pregnancy loss . Exhibit N details the specific codes used to identify VHA-covered obstetric deliveries across the 16-year period. To accommodate changes that the Centers for Medicare & Medicaid Services' (CMS) made to the classification of DRGs in 2008, we applied one set of inclusion/exclusion criteria for identifying obstetric deliveries to the FY00 and FY05 Veteran patient cohorts and a slightly modified definition to the FY10 and FY15 Veteran patient cohorts.

Exhibit N. Inclusion and Exclusion Criteria for Obstetric Deliveries

	Variable(s)	FY00 and FY05	FY10 and FY15
Inclusion Criteria	DXLSF, DX2-DX5	V27xy, 650xy	V27xy, 650xy
	PDRG	370, 371, 372, 373, 374, 375	765, 766, 767, 768, 774, 775
	SURG9CD1-SURG9CD5	720xy, 721xy, 724xy, 726xy, 728xy, 729xy, 736xy, 7221x, 7229x, 7231x, 7239x, 7251x, 7252x, 7253x, 7254x, 7271x, 7279x, 7322x, 7359x, 740xy,741xy, 742xy, 744xy, 7499x	720xy, 721xy, 724xy, 726xy, 728xy, 729xy, 736xy, 7221x, 7229x, 7231x, 7239x, 7251x, 7252x, 7253x, 7254x, 7271x, 7279x, 7322x, 7359x, 740xy,741xy, 742xy, 744xy, 7499x
Exclusion Criteria	DXLSF, DX2-DX5	630xy, 631xy, 632xy, 633xy, 634xy, 635xy, 636xy, 637xy, 638xy, 639xy	630xy, 631xy, 633xy, 632xy, 634xy, 635xy, 636xy, 637xy, 638xy, 639xy
	PDRG	378, 380, 381	770, 777, 779
	SURG9CD1-SURG9CD5	750xy, 6901x, 6951x, 7491x	750xy, 6901x, 6951x, 7491x

NOTE: A small number of U.S. births occur out of the hospital[45] (i.e., at home, in birthing centers, in Emergency Departments, or in other outpatient settings), but only inpatient deliveries are reported in this Sourcebook. This could result in a small underestimation of deliveries. For example, in FY15 there were 366 women Veterans who did not have an obstetric delivery appearing in the inpatient Purchased Care records but who did have evidence of an obstetric delivery in the form of a CPT code[46] appearing in

[43] Kuklina EV, Whiteman MK, Hillis SD, et al. An enhanced method for identifying obstetric deliveries: implications for estimating maternal morbidity. *Matern Child Health J.* Jul 2008;12(4):469-477.
[44] Shaw JG, Asch SM, Kimerling R, Frayne SM, Shaw KA, Phibbs CS. Posttraumatic stress disorder and risk of spontaneous preterm birth. *Obstet Gynecol.* Dec 2014;124(6):1111-1119.
[45] For example, see https://www.cdc.gov/nchs/products/databriefs/db144.htm.
[46] In the outpatient Purchased Care setting, we identified obstetric deliveries by one or more instances of at least one of the following CPT codes: 59400, 59409, 59410, 59510, 59514, 59515, 59610, 59612, 59614, 59618, 59620, 59622.

the outpatient Purchased Care records.[47]

[47] For these 366 deliveries appearing in the outpatient Purchased Care file, the setting where the delivery occurred (identified by the place of service, PLSER, variable) was an inpatient hospital for 87% (318 out of 366). The remaining 48 deliveries occurred in birthing centers, outpatient hospitals, office settings, military treatment facilities, at home, or in the Emergency Department.

6.0 Algorithms for Part 3: Health Profile

6.1. Overview

To characterize the medical and mental health conditions of women Veterans, we used International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis codes appearing in VHA outpatient/inpatient files and in Purchased Care outpatient/inpatient files. In the outpatient setting, the clinician records the ICD-9-CM diagnosis/diagnoses addressed at the visit on an encounter form, which is then incorporated into the patient's administrative records. In the inpatient setting, coders typically abstract admitting diagnoses and discharge diagnoses appearing in the patient's medical record, which are then incorporated as ICD-9-CM codes into the patient's administrative records. To capitalize on these existing clinical data, we aggregated the more than 15,000 ICD-9-CM diagnosis codes into 202 clinically meaningful "***conditions,***" and then grouped conditions into 17 broad "***domains:***" Infectious Disease; Endocrine/Metabolic/Nutritional; Cardiovascular; Respiratory; Gastrointestinal; Urinary; Reproductive Health; Breast; Cancer; Hematologic/Immunologic; Musculoskeletal; Neurologic; Mental Health/SUD; Sense Organ; Dental; Dermatologic; and Other conditions. This section describes the five phases of variable creation completed for Sourcebook Volume 3 with minor updates for Sourcebook Volume 4.[48]

- *Phase* 1 involved developing a rule for mapping ICD-9-CM codes to conditions.
- *Phase 2* involved developing a rule for mapping conditions to domains.
- *Phase 3* involved database processing to generate person-level indicators for presence of each condition within each data source (VHA outpatient, VHA inpatient, Purchased Care outpatient, and Purchased Care inpatient).
- *Phase 4* synthesized information across data sources, applying an algorithm for identification of presence/absence of each condition for each patient.
- *Phase 5* generated patient-level variables indicating, for each domain, whether or not the patient had at least one condition falling within the domain.

Detailed description of these phases follows.

6.2. Phase 1: Rule for Mapping ICD-9-CM Codes to Conditions

There are two major reasons for the decision to map ICD-9-CM codes to broader "conditions". First, attempting to present the frequency of each individual ICD-9-CM diagnosis code would be more confusing than illuminating, as there are well over 15,000 ICD-9-CM diagnosis codes. Second, in many cases a clinician coding the diagnosis responsible for the patient's visit or hospital stay could legitimately apply one of several ICD-9-CM codes to reflect the presenting condition. For example, if the clinician identifies migraine headache as the patient's diagnosis at a visit, then the clinician could code the reason for that visit as ICD-9-CM 346.00 ("migraine with aura, without mention of intractable migraine"), as ICD-9-CM 346.90 ("migraine, unspecified, without mention of intractable migraine"), or as ICD-9-CM 784.0 ("headache"), among other options, all to describe the same clinical presentation. Similarly, a clinician seeing a patient for diabetes mellitus might correctly code the reason for the visit as ICD-9-CM 250.60 ("diabetes type II or unspecified type, with neurological manifestations"), as ICD-9-CM 250.90 ("diabetes type II or unspecified type, with unspecified complication"), or as ICD-9-CM 357.2 ("polyneuropathy in diabetes"), among other options. In other words, *to present data from a single ICD-9-CM diagnosis code may be to apply a higher level of granularity of results than typical clinician coding practices would support*. Therefore, it is necessary to aggregate ICD-9-CM codes into groupings

[48] Specifically, Sourcebook Volume 4 reports only on 1+ instance of a condition or domain; therefore, duplicate processing of the outpatient files and the creation of person-level count variables for each condition or domain was not necessary for this volume.

meaningful to the purpose of the work being pursued.

Fortunately, a widely-used approach to aggregating ICD-9-CM codes exists. The Agency for Healthcare Research and Quality (AHRQ) sponsors the Healthcare Cost and Utilization Project (HCUP) to develop Clinical Classification Software (CCS) that categorizes all ICD-9-CM diagnosis codes into a set of clinically meaningful groups, each reflecting a single condition.[49] As with Sourcebook Volume 3, for Sourcebook Volume 4, the CCS approach serves as the foundation for the WHEI strategy for mapping ICD-9-CM codes to conditions; the multi-step process WHEI used to tailor the CCS approach to the needs of the Sourcebook reports is described next.

In Step 1, WHEI used as a starting point an existing mapping strategy. Specifically, starting with the 2008 version of CCS,[50] a prior research study[51] made modifications to CCS's mapping strategy based upon clinical input, to enhance its suitability for describing burden of illness in women Veteran VHA patients; some ICD-9-CM codes (e.g., pediatric illnesses, congenital conditions, E codes, and codes reflecting a procedure rather than a diagnosis) were not mapped to a condition in that prior work.[52] The 237 conditions (222 medical conditions plus 15 mental health conditions[53]) from this prior work constituted the Step 1 working list of conditions.

In Step 2, Women's Health Services' Cardiovascular Health Workgroup—composed of cardiologists, primary care providers, policy-makers and women's health researchers—carefully reviewed the Step 1 working list and through an iterative group consensus process refined the cardiovascular conditions on that list. Careful development of these conditions was considered a high priority, given that cardiovascular disease remains the leading cause of death for women. Their work is described in a report on women Veterans' cardiovascular health.[54]

In Step 3, Women's Health Services' Reproductive Health Workgroup—composed of gynecologists, primary care providers, policy-makers and women's health researchers—similarly reviewed the Step 1 working list and used an iterative group consensus process, combined with review of American College of Obstetrics & Gynecology coding guidelines,[55] to refine the reproductive health conditions on that list. Careful development of these conditions was considered a priority because female-specific reproductive health conditions are unique to women and require a specialized health care delivery infrastructure, falling under the purview of Women's Health Services. The work of the Reproductive Health Workgroup is described in a reproductive health report.[56]

In Step 4, the WHEI team resolved any overlap between condition lists generated by these two

[49] HCUP Tools and Software. Healthcare Cost and Utilization Project (HCUP). December 2017. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/tools_software.jsp.
[50] Healthcare Cost and Utilization Project (HCUP) Clinical Classification Software (CCS). Agency for Healthcare Research and Quality. Rockville, MD; 2008.
[51] VA HSR&D SHP 08-161 (PI: Rachel Kimerling PhD; Co-PI: Susan Frayne MD, MPH)
[52] Frayne SM, Chiu VY, Iqbal S, et al. Medical care needs of returning veterans with PTSD: their other burden. Journal of General Internal Medicine 2011;26(1):33-39. Also see that manuscript's On-Line Appendix 2 for a complete list of all 222 conditions.
[53] The mapping strategy for mental health conditions used in that study—similarly representing a modified version of the CCS mapping strategy—additionally drew upon mapping algorithms developed for other prior studies: VA HSR&D IIR 04-248, and NIDDK 1 R01 DK071202.
[54] Maher NH, Whitehead AM, Duvernoy C, Davis M, Safdar B, Frayne SM, Saechao F, Lee J, Hayes PM, Haskell SG. The State of Cardiovascular Health in Women Veterans. Volume 2: Risk Factors, Diagnoses, and Procedures in Fiscal Year (FY) 2014. Women's Health Evaluation Initiative, Women's Health Services, Veterans Health Administration, Department of Veterans Affairs, July, 2017.
[55] American College of Obstetricians & Gynecologists. OB/GYN Coding Manual: Components of Correct Procedural Coding. American Medical Association, 2010. Print.
[56] Zephyrin LC, Katon J, Hoggatt KJ, Balasubramanian V, Saechao F, Frayne SM, Mattocks KM, Feibus K, Galvan IV, Hickman R, Hayes PM, Haskell SG, Yano EM. State of Reproductive Health In Women Veterans - VA Reproductive Health Diagnoses and Organization of Care. Women's Health Services, Veterans Health Administration, Department of Veterans Affairs, Washington, DC. February 2014.

workgroups. For example, "Hypertension in Pregnancy" appeared on both the Cardiovascular Health Workgroup list and on the Reproductive Health Workgroup list. This resulted in a working list of 268 conditions at the end of Step 4.

In Step 5, a VHA primary care women's health provider/researcher (SMF) reviewed all of the CCS 2008 ICD-9-CM codes (and their associated CCS conditions) that had not been incorporated into our Step 1 working list of conditions. (These represented mostly childhood conditions, congenital abnormalities, E codes, and medical procedures that do not actually reflect a diagnosis.) We either mapped these codes to the corresponding CCS condition or, in consultation with VHA women's health clinicians and experts from Steps 2 and 3, we mapped the code to one of our Step 4 conditions. There were 316 conditions at the end of Step 5.

In Step 6, a VHA primary care women's health provider/researcher (SMF) reviewed all of the CCS 2012 ICD-9-CM codes (and their associated CCS conditions) that had not been mapped by the end of Step 5. (These represented new ICD-9-CM codes that had not been in existence in 2008.) As in Step 5, our default was to map these ICD-9-CM codes to the corresponding CCS condition, but the clinical reviewer queried other clinicians with relevant expertise for any potentially controversial mappings.

In Step 7, a VHA primary care women's health provider/researcher (SMF) conducted a global consistency review of the ICD-9-CM codes mapped to each condition as of the end of Step 6. This review was intended to confirm that our mapping had yielded clinically meaningful groups of ICD-9-CM codes, and that the grouping logic was consistent across conditions—if not, any necessary adjustments were made. The review also identified conditions that needed to be "lumped" with other conditions and conditions that needed to be "split" into more than one condition. To facilitate lumping and splitting, the reviewer referred to existing nosologies (such as the CCS nosology, the ICD-9-CM nosology, and the nosologies defined by the table of contents structure of major medical textbooks). To further guide this process and inform decisions about lumping versus splitting, for every ICD-9-CM code, we also determined the number of women who had at least one instance of that code in FY12 in the VHA outpatient file, in the VHA inpatient files, in the Purchased Care outpatient file, and in the Purchased Care inpatient file. The reviewer additionally consulted these ICD-9-CM code-level frequencies when making decisions about lumping or splitting: ICD-9-CM codes with particularly high frequency were inspected to assess whether they merited inclusion in a new condition, and conversely if a condition contained only very low frequency ICD-9-CM codes, the condition was assessed for possible lumping with another condition. The reviewer also considered the codes numerically close to the ICD-9-CM code being examined, assessing whether or not the code should be grouped with codes close to it in the ICD-9-CM hierarchy. For difficult grouping decisions, the reviewer also consulted with clinical experts, textbooks, and other reference materials.

In Step 8, we checked our mapping against other sources. Our intention was to make a particular effort to maximize consistency of our approach to condition mapping with the approaches used by other national VHA program offices and VHA Quality Enhancement Research Initiatives (QUERI), to the extent possible. On occasion, this was not possible. For example, in some cases the algorithm of one office differed from that of another office—in that case we attempted to rely on the algorithm of the office primarily responsible for oversight of care for that condition. Likewise, in some cases the algorithm of an office involved mapping a single ICD-9-CM code to more than one condition—because our approach, like that of CCS, required that an ICD-9-CM code be mapped uniquely to a single condition, we converted such double-mappings to single-mappings. We received detailed condition specification algorithms from the following: VA Mental Health's Northeast Program Evaluation Center (NEPEC) (for a number of the

major mental health conditions examined in this Sourcebook); the Military Sexual Trauma Support Team, the Program Evaluation Resource Center (PERC), and the Substance Use Disorder QUERI (for additional checks of mental health and substance use disorder conditions); Office of Geriatrics and Extended Care (for dementia); Polytrauma and Blast-Related Injuries QUERI (for Traumatic Brain Injury, TBI); Chronic Heart Failure QUERI (for heart failure) and Spinal Cord Injury QUERI (for Spinal Cord Injury). We also reviewed key publications of the Diabetes QUERI (for diabetes mellitus[57]) and the Stroke QUERI (for stroke[58]). Several investigators also generously shared their ICD-9-CM mapping algorithms for specific conditions with us. Cross-checking against these various sources led to final adjustments to our algorithm for mapping ICD-9-CM codes to conditions, yielding 234 conditions at the end of Step 8.

In Step 9, we pruned the list of conditions to include only those relevant to the Sourcebook. While the final list of conditions available at the end of Step 8 mapped every ICD-9-CM code to a condition, not all of these conditions are of interest for this report. We deleted "conditions" representing E codes because they describe a mechanism of injury or type of accident rather than a condition—an additional ICD-9-CM code should accompany the E code to specify the type of injury sustained, and that is the code retained for this report. We deleted "conditions" representing neonatal codes because our focus is upon the mother and not the newborn. We also deleted "conditions" that did not actually represent presence of a medical condition, many of which were composed of V codes.[59] For example, in general, we deleted "conditions" reflecting receipt of a medical procedure; these conditions contained ICD-9-CM codes such as V70.0 "routine general medical examination at a health care facility" or V54.89 "other orthopedic aftercare" or V76.10 "breast screening unspecified" or V57.1 "care involving other physical therapy". One exception to this approach is that we did count contraceptive counseling (which is technically a medical preventive/counseling code rather than a clinical diagnosis) as one of the conditions reported in this Sourcebook, because of its particular relevance to the medical care of women. We also deleted other "conditions" not indicating presence of a medical condition, such as conditions containing ICD-9-CM codes describing family history (e.g., V17.1 "family history of stroke [cerebrovascular]") or non-disease clinical observations (e.g., 795.5 "nonspecific reaction to tuberculin skin test without active tuberculosis").

Sourcebook Volume 4 reports on a total of 12,912 ICD-9-CM codes mapped to a total of ***202 conditions*** (unchanged from Sourcebook Volume 3). These 202 conditions are presented in this Sourcebook.

See Online Appendix (Supplemental Materials) for the complete mapping of ICD-9-CM codes to conditions for FY15.

6.3. Phase 2: Rule for Mapping Conditions to Domains

Applying clinical expertise and drawing upon the broad groupings developed by CCS, a panel of VHA women's health primary care providers and researchers grouped these 202 conditions into 17 broad

[57] Miller DR, Safford MM, Pogach LM. Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. *Diabetes Care* 2004;27(Suppl 2):B10-21.
[58] Reker DM, Rosen AK, Hoenig H, et al. The hazards of stroke case selection using administrative data. *Med Care* 2002;40(2):96-104.
[59] While the majority of V codes were mapped to conditions not included in this Sourcebook, some were mapped to conditions that were included, because these ICD-9-CM codes indicated presence of a medical condition. Illustrative examples include: V07.4 "hormone replacement therapy (postmenopausal)" was mapped to Menopausal Disorders; V10.05 "personal history of malignant neoplasm of large intestine" was mapped to Cancer - Colorectal; V15.41 "personal history of physical abuse" was mapped to Psychosocial Factors – Other; V22.1 "supervision of other normal pregnancy" was mapped to Pregnancy or Delivery - Normal; V25.09 "other general counseling and advice on contraceptive management" was mapped to Contraceptive Care Management; V41.2 "problems with hearing" was mapped to Hearing Problems; V43.3 "heart valve replaced by other means" was mapped to Valvular Disease; V58.11 "encounter for antineoplastic chemotherapy" was mapped to Cancer - Other and Unspecified Primary; V60.0 "lack of housing" was mapped to Housing Insufficiency; V62.82 "bereavement uncomplicated" was mapped to Psychosocial Factors – Other; V69.0 "lack of physical exercise" was mapped to Residual Codes; V85.35 "body mass index 35.0-35.9, adult" was mapped to Overweight/Obesity.

"domains" that primarily represent organ systems. Each condition received a primary mapping to a single domain. Primary mappings are reflected in Part 3, Exhibit 3.F, which lists every condition categorized by its primary domain. Some conditions also were secondarily mapped to another domain; in that case, the condition was counted both toward the frequency of the primary domain and toward the frequency of the secondary domain. The conditions with secondary mappings are listed below, by domain (with their primary domain mapping in parentheses). The 17 domains are as follows.

1. **Infectious Disease domain.** Systemic infections and unspecified infections receive primary mapping to Infectious Disease. Infections of a specific organ system are primarily mapped to that organ system, and secondarily mapped to Infectious Disease. For example, the condition "Hepatitis C" is primarily mapped to the Gastrointestinal domain, and secondarily mapped to the Infectious Disease domain.[60] The conditions secondarily mapped to Infectious Disease for total Infectious Diseases counts were the following:
   - Pneumonia (primary domain: Respiratory)
   - Respiratory System Infections - Other (primary domain: Respiratory)
   - Hepatitis C (primary domain: Gastrointestinal)
   - Urinary Tract Infection (Cystitis/Urethritis/Pyelonephritis) (primary domain: Urinary)
   - Sexually Transmitted Infections (primary domain: Reproductive Health)
   - Vaginitis and Other Pelvic Inflammatory Conditions (primary domain: Reproductive Health)
   - Osteomyelitis/Infectious Arthritis (primary domain: Musculoskeletal)
   - Skin Infection (primary domain: Dermatologic)
2. **Endocrine/Metabolic/Nutritional domain.** Endocrine, metabolic and nutritional disorders are primarily mapped to this domain. Conditions secondarily mapped to this domain were the following:
   - Pregnancy Complicated by Diabetes Mellitus (primary domain: Reproductive Health)
   - Cancer – Thyroid (primary domain: Cancer)
3. **Cardiovascular domain.** This refers to conditions that affect the heart and other parts of the cardiovascular system, including cerebrovascular and peripheral vascular conditions. One condition was secondarily mapped to this domain.
   - Pregnancy Complicated by Hypertension (primary domain: Reproductive Health)
4. **Respiratory domain.** This includes conditions that affect the lungs and upper respiratory tract. One condition was secondarily mapped to this domain.
   - Cancer – Bronchopulmonary (primary domain: Cancer)
5. **Gastrointestinal domain.** This refers to conditions that affect the digestive system. Conditions secondarily mapped to this domain are as follows:
   - Cancer – Esophagus (primary domain: Cancer)
   - Cancer – Gastric (primary domain: Cancer)
   - Cancer – Colorectal (primary domain: Cancer)
   - Cancer – Anal (primary domain: Cancer)
   - Cancer – Hepatobiliary (primary domain: Cancer)
   - Cancer – Pancreatic (primary domain: Cancer)
6. **Urinary domain.** This refers to conditions of the kidneys, bladder, or other parts of the urinary system. Conditions secondarily mapped to this domain are as follows:
   - Cancer – Renal (primary domain: Cancer)

[60] In a limited number of instances (specified here), conditions were counted toward the total domain count of more than one domain. However, individual ICD-9-CM codes were not counted toward more than one condition.

- Cancer – Bladder (primary domain: Cancer)

7. **Reproductive Health domain.** This encompasses genital tract conditions, pregnancy-related conditions, and other conditions related to reproductive health. Conditions secondarily mapped to this domain are as follows:
   - Cancer – Cervical (primary domain: Cancer)
   - Cancer – Uterine (primary domain: Cancer)
   - Cancer – Ovarian (primary domain: Cancer)
   - Cancer – Female Reproductive – Other (primary domain: Cancer)
   - Carcinoma in Situ – Cervical (primary domain: Cancer)
   - Carcinoma in Situ – Female Reproductive – Other (primary domain: Cancer)
   - Cancer – Prostate (primary domain: Cancer)
   - Cancer – Testicular (primary domain: Cancer)
8. **Breast domain.** This includes breast conditions and abnormal breast findings. Conditions secondarily mapped to this domain are as follows:
   - Cancer – Breast (primary domain: Cancer)
   - Carcinoma in Situ – Breast, Ductal or Lobular (primary domain: Cancer)
9. **Cancer domain.** All cancer diagnoses and all carcinoma in situ diagnoses[61] are primarily mapped to the Cancer domain. Whenever applicable, cancers are secondarily mapped to the organ system to which they refer.
10. **Hematologic/Immunologic domain.** This refers to disorders of the blood or immune system. Conditions secondarily mapped to this domain are as follows:
    - Lymphomas (primary domain: Cancer)
    - Leukemias (primary domain: Cancer)
    - Multiple Myeloma (primary domain: Cancer)
11. **Musculoskeletal domain.** This includes rheumatologic and musculoskeletal conditions. One condition is secondarily mapped to this domain.
    - Cancer – Bone/Connective Tissue (primary domain: Cancer)
12. **Neurologic domain.** This refers to conditions of the brain and nervous system. Conditions secondarily mapped to this domain are as follows:
    - Cerebrovascular Accident/Transient Ischemic Attack (primary domain: Cardiovascular)
    - Cancer – Brain/Nervous System (primary domain: Cancer)
13. **Mental Health/SUD domain.** This domain consists of mental health conditions, SUDs, and nonspecific psychiatric disorders. Note that general psychosocial factors and Tobacco Use Disorder are mapped to the Other domain, not to the Mental Health/SUD domain.
14. **Sense Organs domain.** This includes conditions that affect the eyes or ears.
15. **Dental domain.** This refers to dental disorders.

    NOTE: Most Veteran VHA patients are not eligible to receive dental care by a VHA provider; therefore frequencies of dental disorders among Veteran VHA patients may represent an under-count of true condition prevalence.

16. **Dermatologic domain.** This refers to conditions affecting the skin. One condition is secondarily mapped to this domain.

[61] Carcinoma in Situ can in some cases represent a condition managed as cancer (e.g., ductal breast carcinoma in situ) and in other cases can represent a non-cancer condition (e.g., cervical carcinoma in situ). However, because the latter is serious and on the pathway toward cancer, all Carcinoma in Situ is grouped within the Cancer domain.

- Melanoma (primary domain: Cancer)

17. **Other domain.** This domain includes miscellaneous diagnoses not mapped to other domains, such as symptoms, conditions due to external causes, and psychosocial factors.

6.4. Phase 3: Generating Person-Level Variables for Each Condition Within Each Data Source

In this phase, we processed raw record-level data from the four source files (VHA outpatient, VHA inpatient, Purchased Care outpatient, and Purchased Care inpatient files), with a goal of generating person-level indicator variables for each of the 202 conditions for each of the source files in each fiscal year. For example, in the case of the condition "Diabetes Mellitus," the goal of Phase 3 was to create four variables, as follows:

1. From the VHA outpatient file: a person-level *indicator variable (yes/no)* indicating whether at least one instance of a Diabetes Mellitus ICD-9-CM code appeared in the file;
2. From the VHA inpatient file: a person-level *indicator variable (yes/no)* indicating whether at least one instance of a Diabetes Mellitus ICD-9-CM code appeared in the file;
3. From the Purchased Care outpatient file: a person-level *indicator variable (yes/no)* indicating whether at least one instance of a Diabetes Mellitus ICD-9-CM code appeared in the file; and
4. From the Purchased Care inpatient file: a person-level *indicator variable (yes/no)* indicating whether at least one instance of a Diabetes Mellitus ICD-9-CM code appeared in the file.

The source files used for database processing were:

1. VHA Outpatient Event files
2. VHA Inpatient Main, Bedsection and Census files
3. Purchased Care Outpatient Services files
4. Purchased Care Inpatient Stay and Ancillary files

Creating these person-level variables involved modifying the raw record-level files as described next.

Step 1: Exclude outpatient records not representing a face-to-face encounter with a clinician. In the outpatient files,[62] we excluded records that did not represent a face-to-face encounter with a clinician (based upon clinic stop codes for VHA outpatient records and CPT codes for Purchased Care outpatient records[63]), such as Laboratory encounters[64] and most Radiology encounters,[65,66] telephone encounters, Store & Forward encounters, Home Telehealth encounters, and Secure Messaging encounters. The resulting outpatient files contained only records for face-to-face encounters with a clinician, i.e., settings in which a diagnosis can legitimately be made by a clinician.

Step 2: Elongate the record-level file. To address the fact that a single utilization record may contain more than one ICD-9-CM diagnosis field (and that the number of diagnosis fields differs in different source files), we created an elongated file with one non-missing ICD-9-CM diagnosis per record. For instance, a single record with 10 diagnoses in the raw data file was elongated into 10 records, each with a single diagnosis, while all of the other fields remained constant.

[62] Step 1 was not necessary for inpatient files, because all inpatient stays are considered face-to-face encounters with a clinician.
[63] See Online Appendix (Supplemental Materials) for the complete list of FY15 clinic stop codes that were considered face-to-face encounters in the VHA outpatient setting and the FY15 CPT codes that were considered face-to-face services in the Purchased Care outpatient setting.
[64] Among all records with a laboratory clinic stop code in the FY15 VHA outpatient file, 89% were missing ICD-9-CM codes across all diagnosis fields.
[65] Among all records with a radiology clinic stop in the FY15 VHA outpatient file, 66% were missing ICD-9-CM codes across all diagnosis fields.
[66] Some Radiology encounters were considered face-to-face encounters with a clinician, such as invasive radiology procedures that require a Radiologist evaluation (and thus a clinical diagnosis) prior to performing the procedure.

This step was completed for the VHA outpatient and inpatient files and the Purchased Care inpatient files. The Purchased Care outpatient files have only one diagnosis per record.[67]

Step 3: Create person-level variables for each condition *within* data sources. For each of the 202 conditions we generated four person-level variables indicating whether an ICD-9-CM code mapping to the condition appeared at least once in any VHA outpatient record (yes/no), at least once in any VHA inpatient record (yes/no), at least once in any Purchased Care outpatient record (yes/no), and at least once in any Purchased Care inpatient record (yes/no).

6.5. Phase 4: Generating Final Person-Level Variables for Each Condition Across Data Sources

To generate the final person-level variable for each condition, we created an additional across-file condition indicator variable (yes/no for presence of the condition) that synthesized information from the within-file person-level condition variables.

All analyses presented in this Sourcebook use the following *Base Algorithm* for conditions.[68]

A patient is considered to have a particular condition in FY15 if she/he has **at least one instance of an ICD-9-CM code mapped to the condition in FY15 in an outpatient record (VHA or Purchased Care files, limited to face-to-face visits with a clinician) or in an inpatient record (VHA or Purchased Care files).**

6.6. Phase 5: Generating Final Person-Level Variables for Each Domain

Finally, using the Base Algorithm person-level condition variables generated in Phase 4, and applying the mapping strategy described in Phase 2, for each of the 17 domains we created a person-level indicator variable (yes/no) indicating whether the patient had at least one condition falling within that domain.

6.7. Calculating Age-Adjusted Odds Ratios

For each domain and for each condition individually, we calculated age-adjusted odds ratios that compared the odds of the condition in women compared to men. For these analyses, age was treated as an ordinal variable: 18-39 (reference group), 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+.

[67] In Sourcebook Volume 3, we described an additional Step 3 in which we processed duplicates in the outpatient data. It was not necessary to process duplicates or overlapping admissions in VHA inpatient files nor in Purchased Care inpatient files because we were only interested in whether there was *at least one instance* of the condition in the fiscal year (i.e., we did not need to generate count variables for inpatient conditions).

[68] In Sourcebook Volume 3, we additionally created a High Sensitivity algorithm (presence of at least one instance of the condition across data sources in FY11-FY12) and a High Specificity Algorithm (presence of at least one instance of the condition in FY12 AND at least one confirmatory instance of the condition in FY11-FY12 outpatient files, OR presence of at least one instance of the condition in the FY12 inpatient files). See Sourcebook Volume 3, Online Appendix D. Comparison of domain frequencies among women Veteran patients using Base, High Sensitivity and High Specificity Algorithms, FY12. Available at: https://www.womenshealth.va.gov/WOMENSHEALTH/sourcebookvol3onlineappendix.asp.

7.0 Algorithm for Part 4: Geographic Distribution

7.1. Overview

A major new component of Sourcebook Volume 4 is characterization of the geographic distribution of women Veterans across the United States and its territories over time. As of the start of FY15, VHA was organized into 21 distinct Veterans Service Integrated Networks, or VISNs, that represent broad geographic administrative areas of the United States.[69] Each VISN is comprised of multiple Health Care Systems, which usually contain a flagship VA Medical Center and several Community-Based Outpatient Clinics. This Sourcebook examines the distribution of women Veterans at both the VISN-level and the Health Care System-level.

The geographic distribution variables are derived from the Medical SAS VHA Outpatient Event (SE)[70] files. Next, we describe the three phases involved in creating the geographic distribution variables.

7.2. Phase 1: Mapping of VISNs Over Time

We compared the list of unique VISNs that appeared in the FY00 SE file with the list of unique VISNs that appeared in the FY15 SE file. The list of VISNs was the same in FY00 and FY15, with one exception: VISN 13 merged with VISN 14 to create a new VISN called VISN 23 in FY12. For comparability between FY00 and FY15, Sourcebook Volume 4 combines VISN 13 and VISN 14 data when reporting the number of women Veterans per VISN in FY00.

7.3. Phase 2: Mapping of Health Care Systems Over Time

Over the 16-year period, VHA expanded to better accommodate the needs of Veterans and provide greater access for Veterans. Some of the expansion efforts involved the opening of new Health Care Systems, merging of VHA facilities into other Health Care Systems, and closure of Health Care Systems.[71]

To be able to compare Health Care System-level changes in patient populations from FY00 to FY15, we needed to create a master list of Health Care Systems that could be applied to both FY00 data and FY15 data. As a first step, we created the FY15 WHEI master list of Health Care Systems to match the FY15Q4 VAST list of 141 Health Care Systems, except that the WHEI master list split the New York Harbor Health Care System (station 630) into two separate Health Care Systems: Manhattan Division (station 630) and Brooklyn Division (station 630A4).[72]

These changes resulted in 142 Health Care Systems in FY15.

To create the FY00 WHEI master list of Health Care Systems, we started with the FY15 WHEI master list of Health Care Systems, since the VAST list was not available prior to FY09, and compared it to the Health Care Systems identified from the STA3N field of the FY00 SE file. Three Health Care Systems

[69] During 2015, VA initiated some realignment of VISN boundaries so that all VISNs would be contained within state lines. There are three exceptions: California is split between VISN 21 and VISN 22, Illinois is split between VISN 12 and VISN 15, and Michigan is split between VISN 10 and VISN 12. Because these changes were not fully implemented by the end of FY15, the analyses presented in this volume do not reflect these changes.

[70] In the SE file, Health Care Systems, or parent stations, are usually identified from the 3-digit STA3N variable, while the individual site of care is identified from the 5-digit STA5A variable (though parent stations also appear in the STA5A field as a 3-digit station number). There are also some Health Care Systems that appear as 5-digit stations; See Section 7.3. Phase 2: Mapping of Health Care Systems over time.

[71] Additionally, some Community-Based Outpatient Clinics may have opened, closed, or merged with other Community-Based Outpatient Clinics or with other Health Care Systems during the 16-year time period from FY00-FY15.

[72] The Women's Health Assessment Tool for Comprehensive Health (WATCH) survey, conducted annually by Women's Health Services, considers Manhattan and Brooklyn to be two separate Health Care Systems. A Women Veteran Program Manager is assigned to each site.

were present in the FY00 SE file but were not present in FY15 due to merges with other Health Care Systems: Lincoln NE, Fort Lyon CO, and Murfreesboro, TN. For comparability between FY00 and FY15, Lincoln (station 597) was mapped to Omaha (station 636), Fort Lyon (station 567) was mapped to Denver (station 554), and Murfreesboro (station 622) was mapped to Nashville (station 626).

The FY00 WHEI master list contained 140 Health Care Systems. There were two fewer Health Care Systems in the FY00 WHEI master list than in the FY15 WHEI master list because two new facilities (which opened prior to FY15) were not present in FY00: East Central Florida Health Care System (Lake Nona, Orlando VA Medical Center) and VA Texas Valley Coastal Bend Health Care System (Harlingen).

Exhibit O provides the final mapping of Health Care Systems to VISNs.

Exhibit O. Mapping of Health Care Systems and VISNs

VISN (SE)	Health Care System Mapping		Common Name (VAST)
	FY00 Station (WHEI)	FY15 Station (WHEI)	
1	402	402	Togus
1	405	405	White River Junction
1	518	518	Bedford
1	523	523	Jamaica Plain
1	608	608	Manchester
1	631	631	Central Western Massachusetts
1	650	650	Providence
1	689	689	West Haven
2	528	528	Buffalo
2	528A5	528A5	Canandaigua
2	528A6	528A6	Bath
2	528A7	528A7	Syracuse
2	528A8	528A8	Albany
3	526	526	Bronx
3	561	561	East Orange
3	620	620	Montrose
3	630	630	Manhattan
3	630A4	630A4[(a)]	Brooklyn
3	632	632	Northport
4	460	460	Wilmington
4	503	503	Altoona
4	529	529	Butler
4	540	540	Clarksburg
4	542	542	Coatesville
4	562	562	Erie
4	595	595	Lebanon
4	642	642	Philadelphia
4	646	646	Pittsburgh-University Drive
4	693	693	Wilkes-Barre
5	512	512	Baltimore
5	613	613	Martinsburg
5	688	688	Washington
6	517	517	Beckley
6	558	558	Durham
6	565	565	Fayetteville
6	590	590	Hampton
6	637	637	Asheville
6	652	652	Richmond
6	658	658	Salem
6	659	659	Salisbury
7	508	508	Atlanta-Decatur
7	509	509	Augusta Downtown
7	521	521	Birmingham
7	534	534	Charleston
7	544	544	Columbia
7	557	557	Dublin

VISN (SE)	Health Care System Mapping		Common Name (VAST)
	FY00 Station (WHEI)	FY15 Station (WHEI)	
7	619	619	Montgomery
7	679	679	Tuscaloosa
8	516	516	Bay Pines
8	546	546	Miami
8	548	548	West Palm Beach
8	573	573	Gainesville
8	672	672	San Juan
8	673	673	Tampa
8		675[(b)]	Lake Nona
9	581	581	Huntington
9	596	596	Lexington-Leestown
9	603	603	Louisville
9	614	614	Memphis
9	621	621	Mountain Home
9	626[(c)]	626	Nashville
10	538	538	Chillicothe
10	539	539	Cincinnati
10	541	541	Cleveland
10	552	552	Dayton
10	757	757	Columbus
11	506	506	Ann Arbor
11	515	515	Battle Creek
11	550	550	Danville
11	553	553	Detroit
11	583	583	Indianapolis
11	610	610	Marion
11	655	655	Saginaw
12	537	537	Chicago
12	556	556	North Chicago
12	578	578	Hines
12	585	585	Iron Mountain
12	607	607	Madison
12	676	676	Tomah
12	695	695	Milwaukee
15	589	589	Kansas City
15	589A4	589A4	Columbia
15	589A5	589A5	Topeka
15	589A7	589A7	Wichita
15	657	657	St. Louis-John Cochran
15	657A4	657A4	Poplar Bluff
15	657A5	657A5	Marion
16	502	502	Alexandria-Pineville
16	520	520	Biloxi
16	564	564	Fayetteville
16	580	580	Houston
16	586	586	Jackson
16	598	598	Little Rock
16	623	623	Muskogee

VISN (SE)	Health Care System Mapping		Common Name (VAST)
	FY00 Station (WHEI)	FY15 Station (WHEI)	
16	629	629	New Orleans
16	635	635	Oklahoma City
16	667	667	Shreveport
17	549	549	Dallas
17	671	671	San Antonio
17	674	674	Temple
17		740[b]	Harlingen
18	501	501	Albuquerque
18	504	504	Amarillo
18	519	519	Big Spring
18	644	644	Phoenix
18	649	649	Prescott
18	678	678	Tucson
18	756	756	El Paso
19	436	436	Fort Harrison
19	442	442	Cheyenne
19	554[c]	554	Denver
19	575	575	Grand Junction
19	660	660	Salt Lake City
19	666	666	Sheridan
20	463	463	Anchorage
20	531	531	Boise
20	648	648	Portland
20	653	653	Roseburg
20	663	663	Seattle
20	668	668	Spokane
20	687	687	Walla Walla
20	692	692	White City
21	358	358	Manila-Pasay City
21	459	459	Honolulu
21	570	570	Fresno
21	612	612	Martinez CLC
21	640	640	Palo Alto
21	654	654	Reno
21	662	662	San Francisco
22	593	593	Las Vegas
22	600	600	Long Beach
22	605	605	Loma Linda
22	664	664	San Diego
22	691	691	West Los Angeles
23	437	437	Fargo
23	438	438	Sioux Falls
23	568	568	Fort Meade
23	618	618	Minneapolis
23	636[c]	636	Omaha
23	636A6	636A6	Des Moines
23	636A8	636A8	Iowa City

VISN (SE)	Health Care System Mapping		Common Name (VAST)
	FY00 Station (WHEI)	FY15 Station (WHEI)	
23	656	656	St. Cloud

(a) VAST FY15Q4 combines station 630 and station 630A4 into a single station; the WHEI master list splits these into two stations.
(b) There were two fewer Health Care Systems in the FY00 WHEI master list than in the FY15 WHEI master list because two new facilities (which opened prior to FY15) were not present in FY00: East Central Florida Health Care System (Lake Nona, Orlando VA Medical Center) and VA Texas Valley Coastal Bend Health Care System (Harlingen).
(c) For comparability between FY00 and FY15, Lincoln (station 597) was mapped to Omaha (station 636), Fort Lyon (station 567) was mapped to Denver (station 554), and Murfreesboro (station 622) was mapped to Nashville (station 626).

7.4. Phase 3: Generating Variables for Geographic Distribution

Using the Base Cohorts described in Section 2.3, we generated a count of unique women Veteran patients who had any outpatient encounter at a particular Health Care System in each fiscal year.

We then generated a count of unique women Veteran patients who had any outpatient encounter in a particular VISN in each fiscal year.

NOTE: Unlike the algorithms in the previous sections in which the denominator was VHA patients, the denominator for this section is limited to women Veteran VHA outpatients only.

NOTE: This approach captures the sites where women Veterans received any VHA outpatient care. An individual woman may have contributed towards the count for multiple Health Care Systems, if she received VHA outpatient care from more than one Health Care System. An individual woman may have contributed towards the count for multiple VISNs, if she received VHA outpatient care from more than one VISN.

This Sourcebook reports the geographic distribution of women Veterans in FY00 and FY15. See Online Appendix (Supplemental Materials) for the geographic distribution of women Veterans in FY05 and FY10.