



Guidebook

Guide to the PSSG Enrollee File

Ciaran S. Phibbs, Elizabeth H. Cowgill, Angela Y. Fan

February 2013

Guide to the PSSG Enrollee File. Guidebook.

Health Economics Resource Center (HERC)

VA Palo Alto Health Care System

795 Willow Road (152 MPD)

Menlo Park, CA 94025

Phone: 650-617-2630

Fax: 650-716-2639

Email: herc@va.gov

Suggested citation:

Phibbs CS, Cowgill EH, Fan AY. Guide to the PSSG Enrollee File. Guidebook. Menlo Park, CA. VA Palo Alto, Health Economics Resource Center; 2013 Feb.

Acknowledgements: We gratefully acknowledge contributions from Kevin Willis, Supervisory Geographer at the Planning Systems Support Group (PSSG); Randy Remmel, previous Director of Health Systems Data & Analysis; and Chase Gatlin, Geographer at PSSG. We would also like to acknowledge the previous efforts of Anupama Torgal from the Women’s Health Evaluation Initiative/Women Veterans Health Strategic Health Care Group.

Contents

Terms	v
1. Introduction.....	1
2. Access	2
2.1 Dataset access	2
2.2 Location	2
2.2 More information	2
3. Data Sources	3
3.1 ADUSH Enrollment File.....	3
3.2 Geocoding and zip codes	3
3.3 Urban/rural/highly rural classifications	4
4. Variables	6
4.1 Overview.....	6
4.2 Description of variables	8
References.....	11

Tables and Appendix

Table 1: Variable name and label availability by year (FY07Q4-FY11Q2)	6
---	---

Terms

ADUSH	VHA Assistant Deputy Under Secretary for Health
AITC	Austin Information Technology Center
BIRLS	Beneficiary Identification and Records Locator System
FIPS	Federal information processing standard
FY	Federal fiscal year
GIS	Geographic information system
NCOA	U.S. Postal Service National Change of Address file
NDS	National Data Systems
PSSG	Planning Systems Support Group
SCRSSN	Scrambled social security number
UA	Urbanized area
UC	Urban cluster
URH	Urban/rural/highly rural classification
VA	U.S. Department of Veterans Affairs
VHA	Veterans Health Administration
VISN	Veterans Integrated Services Network
VIReC	VA Information Resource Center
VSSC	VISN Support Services Center

1. Introduction

The U.S. Department of Veterans Affairs (VA) provides health care services to its veterans across the U.S.A., including territories and possessions. Health care services are delivered through 21 geographically divided administrative areas called Veterans Integrated Services Networks (VISN). Over 9 million enrolled veterans and patients receive health services throughout the 21 VISNs.

The Planning Systems Support Group (PSSG) Enrollee File is a geographic dataset that presents the most current study of veterans' health care access, travel, and demand for health services.¹ The Enrollee File is a geocoded file that is produced at the end of each year. The Enrollee File contains a record for each veteran who enrolled to receive health care benefits from the VA since October 1, 1998, including enrollees that may have died since the beginning of the fiscal year. The file also includes those seeking care but not enrolled at the time care was delivered; these are referred to as "cost-only" veterans. The Enrollee File uses the location of residence of each veteran to calculate the travel time and driving distance to the nearest VA provider, for three different levels of VA service (primary, secondary, and tertiary care). The Enrollee File is used by PSSG and the VA Office of Enrollment and Forecasting to geospatially analyze service-level usage, travel time and distance, and health care site forecasting in order to help VISN planners better serve the enrollee population in their network.

The Enrollee File is updated quarterly and at the federal fiscal year-end using information from VA files and the U.S. Postal Service National Change of Address file (NCOA). The Enrollee File also classifies the rurality of enrollee residence using the VA designation as urban, rural, or highly rural (URH) and also includes congressional district, and county/state (FIPS) codes. For over 85% of the records, these variables are based on the veteran's exact residential address; the remaining records assign rural status and travel distance based on the centroid of the residence zip code. Residence address and any other individually identifiable pieces of information are stripped out of the file during processing. The VA scrambled social security number (SCRSSN) is used as a unique identifier and enables researchers to link this data to other VA databases.

This guidebook proceeds as follows: The next section, *Access*, describes where to access to the Enrollee File. The *Data Sources* section describes each of the files used to create the Enrollee File and the methods used to develop the urban/rural/highly rural classification. The guide concludes with the *Variables* section, providing a table of variables within the Enrollee File and a brief description of each variable.

2. Access

This section provides a brief overview on who may access the PSSG Enrollee File, where the dataset can be found, and whom to contact for more information.

2.1 Dataset access

The Enrollee File may be used for studies that have been approved by a Human Subjects review panel with analysts who have permission to use Medical SAS datasets on the Mainframe at the Austin Information Technology Center (AITC). For more information on how to apply for permission to access the Medical SAS datasets and the PSSG Enrollee File, please visit the National Data Systems (NDS) website at <http://vaww4.va.gov/NDS/index.asp>. **Note:** This link directs you to an internal VA website that is not available to the public.

2.2 Location

The PSSG Enrollee File can be found on the AITC Mainframe. The following are the direct location of specific data files:

```
/static/Austin/pssg_geo/PSSG_GEO_FY11Q2sas7bdat  
/static/Austin/pssg_geo/PSSG_GEO_FY11Q1.sas7bdat  
/static/Austin/pssg_geo/PSSG_GEO_FY10Q4.sas7bdat  
/static/Austin/pssg_geo/PSSG_GEO_FY10Q3.sas7bdat  
/static/Austin/pssg_geo/PSSG_GEO_FY10Q2.sas7bdat  
/static/Austin/pssg_geo/PSSG_GEO_FY10Q1.sas7bdat  
/static/Austin/pssg_geo/PSSG_GEO_FY09Q4.sas7bdat  
/static/Austin/pssg_geo/PSSG_GEO_FY08Q4.sas7bdat
```

2.2 More information

For more information, please contact the Planning Systems Support Group (PSSG). PSSG provides both mapping and analytical Geographic Information Systems (GIS) assistance to the Office of the Assistant Deputy Under Secretary for Health and Policy Planning.

Planning Support Systems Group (PSSG)

300 East University Avenue, Suite 430

Gainesville, FL 32601

Phone: (352) 374-6080

Fax: (352) 374-6119

Website: <http://vaww.pssg.med.va.gov/>

Note: This link takes you to an internal VA Intranet website that is not available to the public.

3. Data Sources

This section presents a brief description of the how the PSSG Enrollee File was created.

3.1 ADUSH Enrollment File

The Assistant Deputy Under Secretary for Health (ADUSH) Enrollment File is a large administrative dataset that aggregates information from the following sources: enrollment information from the Veterans Health Administration (VHA) Health Eligibility Center, cost data from the VHA Allocation Resource Center, utilization data from the VHA VISN Support Services Center (VSSC), and death information from the Beneficiary Identification and Records Locator System (BIRLS) Death file and the SSA Death Master file.² The ADUSH Enrollment File contains all enrolled veterans and beneficiary users of VHA services. Veterans are considered to be enrolled if they receive medical care from a VHA service. Each veteran has a single entry for each year but may have multiple records in the cumulative ADUSH Enrollment File. Newly enrolled veterans that have been entered into the ADUSH Enrollment file will also be added into the PSSG Enrollee File each quarter.

The ADUSH Enrollment File is updated with new veterans monthly and cumulatively through the fiscal year-end, and veterans stay in the ADUSH Enrollment File until they are deceased or become ineligible. The ADUSH Enrollment File is updated at the end of each fiscal year for deceased or ineligible veterans. Deceased or ineligible veterans are removed based on their date of death or date of ineligibility. Deceased veterans will not be removed from the ADUSH Enrollment File if there are ongoing costs associated with their name after their death (e.g. cost of fee care that is resolved in the fiscal year after their death) but will be removed once the costs have been resolved. Veterans will be removed from the Enrollee File at the start of the following fiscal year for the same reasons as the ADUSH Enrollment File: death and ineligibility.

Zip code updates to the ADUSH Enrollment File are from three sources: Zip_enrl and Zip_geo files from VHA Health Eligibility Center's National Enrollment Database, and Ziparc from the VHA Allocation Resource Center. The Zip_enrl file is used to update the PSSG Enrollee File.

3.2 Geocoding and zip codes

Geocoding is the process of determining geographical position with latitude and longitude, based on normalized address data.³ For geocoding VA data, PSSG determines veterans' addresses from the the Zip_enrl file from the ADUSH Enrollment File and updated addresses from the U.S. Postal Service National Change of Address file (NCOA). For over 85% of veterans geocoded data will be based on their actual address. For the rest, data will be based on the centroid of the resident zip code.

Geocoded data are grouped by zip code in the PSSG Enrollee File; patient names and addresses are not included. The PSSG Enrollee File provides the most current zip codes and geocoded data

from September 2009 (SEP09) to FY11 Quarter 2 (MAR11).ⁱ Because it is possible for veterans to move several times each year, the dataset is updated each quarter and at the end of each fiscal year. In order to reconcile several zip codes to only one, PSSG uses the most recently reported address.ⁱⁱ

3.3 Urban/rural/highly rural classifications

The urban/rural/highly rural (URH) layer was developed using the following PSSG process.³ First, PSSG defines urban areas, including urbanized areas (UA) and urban clusters (UC), by working with urban area boundary files from the U.S. Census Bureau. UAs and UCs include census blocks or block groups with a population density of at least 1,000 persons per square mile and the surrounding census blocks having a density of at least 500 persons per square mile. Under certain conditions the U.S. Census Bureau designates less densely settled territory as being part of an UA or UC. UCs are typically small towns and named places surrounded by rural or highly rural areas. PSSG determined that it would be prudent to exclude UCs from its definition of urban areas, therefore only UAs are defined as urban areas.

Next, PSSG identifies urban enrollees and facilities. In a GIS process called a "spatial join" the enrollees and facilities located within UAs are assigned the "urban" designation. All other enrollees and facilities are considered "rural" until such time as they are changed to "highly rural."

Then PSSG assigns county federal information processing standard (FIPS) codes. In the same GIS, PSSG performs a spatial join between county shapes (features) and all enrollee and facility point locations. During this process the Census-designated county FIPS code is assigned to each enrollee and facility.

Next, PSSG defines highly rural counties. This step begins with a list of Census-designated highly rural counties and is trimmed to include only those counties with population densities of less than 7 persons per square mile. For reasons not clear at this time, the original Census list of highly rural counties actually includes counties with densities much higher than 7 persons per square mile, even though they exceed the Census Bureau's own defined limit. This is why the PSSG subsets the original list by applying the filter of less than 7 persons per square mile.

Finally, PSSG identifies highly rural enrollees and facilities. Any "rural" enrollee or facility, i.e. one not previously assigned the "urban" designation, with a FIPS code matching the list of highly rural counties is designated "highly rural".

Note: PSSG strongly encourages users to use the URH variable to assign rural status, rather than

ⁱ FY07 and FY08 year-end geocoded information can be found in the AITC file MDPPRD.MDP.SAS.NDE.MAIN.SEPyyGEO.PSSG; however, the process for determining drive times and distance to the closest facility was different before FY08 due to changes in technology.

ⁱⁱ Note: It is possible to do a migration analysis if you merged the 4 quarterly ADUSH Enrollment Files together because each subsequent file gets updated.

zip code, for patients included in the ADUSH Enrollment File. According to PSSG, an address is assigned URH based on whether it is in a metropolitan area as defined by the U.S. Census Bureau and some zip codes include addresses that are in both urban and rural areas.

4. Variables

4.1 Overview

The following table provides a general overview of the variables included in the PSSG Enrollee File and the years which they are present for. An ‘X’ denotes the year where data is available.

Table 1: Variable name and label availability by year (FY07Q4-FY11Q2)

Variable Name	Variable Label	FY 07 Q4 ^a	FY 08 Q4 ^a	FY 09 Q4 ^b	FY 10 Q1 ^b	FY 10 Q2 ^b	FY 10 Q3 ^b	FY 10 Q4 ^b	FY 11 Q1 ^b	FY 11 Q2 ^b
AGEALB	Age last birthday		X ^c	X ^c	X ^d	X ^d	X ^d	X ^e	X ^f	X ^g
CD* (or) CD110**	Congressional district	X*	X**	X*	X**	X**	X**	X*	X*	X*
CLPCSITE	Closest primary care site		X	X				X		
CLSCSITE	Closest secondary care site		X	X				X		
CLTCSITE	Closest tertiary care site		X	X				X		
DDISTPC	Drive distance to closest primary care site		X	X				X		
DDISTSC	Drive distance to closest secondary care site		X	X				X		
DDISTTC	Drive distance to closest tertiary care site		X	X				X		
DTIMEPC	Drive time to closest primary care site		X	X				X		
DTIMESC	Drive time to closest secondary care site		X	X				X		

Variable Name	Variable Label	FY 07 Q4 ^a	FY 08 Q4 ^a	FY 09 Q4 ^b	FY 10 Q1 ^b	FY 10 Q2 ^b	FY 10 Q3 ^b	FY 10 Q4 ^b	FY 11 Q1 ^b	FY 11 Q2 ^b
<u>DTIMETC</u>	Drive time to closest tertiary Site		X	X				X		
<u>FIPS</u>	FIPS code	X	X	X	X	X	X	X	X	X
<u>GENDER</u>	Gender		X	X	X	X	X	X	X	X
<u>LOC_NAME</u>	Geocoding precision indicator							X	X	X
<u>MARKET</u>	Market				X	X	X	X	X	X
<u>PATIENT</u>	Patient flag		X	X	X	X	X	X	X	X
<u>PRIORITY</u>	Adjusted priority		X	X	X	X	X	X	X	X
<u>SCRSSN</u>	Scrambled social security number	X	X	X	X	X	X	X	X	X
<u>SECTOR</u>	Sector	X	X	X	X	X	X	X	X	X
<u>SUBMKT</u>	Submarket				X	X	X	X	X	X
<u>URH</u>	Urban/rural/highly rural flag	X	X	X	X	X	X	X	X	X
<u>VISN</u>	Veterans Integrated Services Network				X	X	X	X	X	X
<u>ZIP_1</u>	Zip code			X	X	X	X	X	X	X
<u>ZIP_2</u>	Zip+4			X						

^a Source: MDPPRD.MDP.SAS.NED.MAIN.SEPyyGEO.PSSG

^b Source: MDPPRD.MDP.SAS.ESR.MAIN.mmmyyGEO.PSSG

^c FY08Q4 and FY09Q4: Per variable label in Proc Contents, age at last birthday as of 9/30/08

^d FY10Q1-FY10Q3: In Proc Contents, variable label does not indicate "AS OF"

^e FY10Q4: Per Variable label in Proc Contents, age at last birthday as of 9/30/10

^f FY11Q1: Per Variable label in Proc Contents, age at last birthday as of 12/31/10

^g FY11Q2: Per Variable label in Proc Contents, age at last birthday as of 3/31/11

4.2 Description of variables

A brief description of all variables in the PSSG Enrollee File included below for quick reference.

AGELB

Age at last birthday.

FY08 Q4 and FY09 Q4: Per variable label in proc contents, age at last birthday as of 09/30/08

FY10 Q1-FY10 Q3: In proc contents, variable label does not indicate “AS OF”

FY10 Q4: Per variable label in proc contents, age at last birthday as of 09/30/10

FY11 Q1: Per variable label in proc contents, age at last birthday as of 12/31/10

FY11 Q2: Per variable label in proc contents, age at last birthday as of 03/31/11

CD* (or) CD110**

Congressional district. Indicates the congressional district in which the veteran lives.

CLPCSITE

Closest primary care site. Provides the station number of the closest primary care site; updated at the end of each fiscal year.

CLSCSITE

Closest secondary care site. Provides the station number of the closest secondary care site; updated at the end of each fiscal year.

CLTCSITE

Closest tertiary care site. Provides the station number of the closest tertiary site; updated at the end of each fiscal year.

DDISTPC

Drive distance to closest primary care site. Computed number of miles to closest primary care site; updated at the end of each fiscal year.

DDISTSC

Drive distance to closest secondary care site. Computed number of miles to closest secondary care site; updated at the end of each fiscal year.

DDISTTC

Drive distance to closest tertiary care site. Computed number of miles to closest tertiary care site; updated at the end of each fiscal year.

DTIMEPC

Drive time to closest primary care site. Drive time in minutes to closest site; updated at the end of each fiscal year.

DTIMESC

Drive time to closest secondary care site. Drive time in minutes to closest secondary care site; updated at the end of each fiscal year.

DTIMETC

Drive time to closest tertiary care site. Drive time in minutes to closest site; updated at the end of each fiscal year.

FIPS

FIPS code. Federal information processing standard (FIPS) codes are a standardized set of codes from the National Institute of Standards and Technology to ensure uniform identification of geographical entities.⁴

GENDER

Gender. Coded as M=Male and F=Female in FY09 Q4.

LOC_NAME

Geocoding precision indicator. Indicates source of geocoding for the following variables: [DTIMEPC](#), [DTIMESC](#), [DTIMETC](#), [FIPS](#), [MARKET](#), [SECTOR](#), [SUBMKT](#), [URH](#), and [VISN](#).

MARKET

Market. Indicates the market in which the veteran lives.

PATIENT

Patient flag. Indicates whether the record is a “user” of VHA.

PRIORITY

Adjusted priority, created in January 2003.

SCRSSN

Scrambled social security number. A formula manipulation of the real social security number and not a randomly generated number. This variable takes a unique value for each real social security number, and thus may be used to identify a patient across fiscal years and datasets.

SECTOR

Sector. Indicates the sector in which the veteran lives.

SUBMKT

Submarket. Indicates the submarket in which the veteran lives.

URH

Urban/rural/highly rural flag. Areas as defined by the U.S. Census Bureau:

Urban area: Census Bureau-defined urbanized area, which is any block or block group having a population density of at least 1000 persons per square mile.

Rural area: Any non-urban or non-highly rural area having between 8 and 999 civilians

per square mile.

Highly rural: Any non-urban area in a county having less than 7 civilians per square mile.

VISN

Veterans Integrated Services Network (VISN). Indicates the VISN in which the veteran lives.

ZIP_1

Zip code. Indicates the zip code in which the veteran lives.

ZIP_2

Zip +4. Indicates the 4 digit extension of the zip code in which the veteran lives.

References

1. Health Economics Resource Center. *New Veteran Geographic Data Source*. VA Health Economics Bulletin. [Newsletter]. 2011 Aug; 11(3). Accessed Jul 2012 at http://www.herc.research.va.gov/files/BLTN_679.pdf.
2. VA Information Resource Center (VIREC). VIREC Research User Guide: 1999-2006 VHA Assistant Deputy Under Secretary for Health (ADUSH) Enrollment File. Accessed Jul 2012 at <http://www.virec.research.va.gov/RUGs/RUG-ADUSH-EF-FY99-06-ER.pdf>.
3. U.S. Department of Veterans Affairs Planning Systems Support Group (PSSG). Utilization. Oct 2012.
4. U.S. Census Bureau. State & County QuickFacts: Federal Information Processing Standards (FIPS). Accessed Mar 2012 at http://quickfacts.census.gov/qfd/meta/long_fips.htm.