Databases for surgical health services research: Veterans Health Administration data

What are Veterans Health Administration(221,480),(312,496) data?

We currently live in the age of “big data.” The data captured within the Veterans Health Administration’s (VHA) electronic health record (EHR) are one of the oldest examples of big data in health services research. The VHA EHR was developed in the mid-1970s and, since its inception, has been managed under the Veterans Information Systems and Technology Architecture (VistA). The current graphical user interface, the Computerized Patient Record System (CPRS), was added in 1997. Each year, VHA data capture health care information for more than 9 million veterans across 170 medical centers and 1,061 community-based outpatient clinics. In 2017 alone, the VHA data included 137 surgery programs that housed 2,065 surgery residents and performed 418,918 surgical procedures.1

The beauty of VHA data is founded in the representation of a complete and nearly unfractured record of health care at both a patient and an organizational level, albeit for a unique population. Regardless of where care was sought in the VHA or which provider treated the patient, all care received within the VHA is mandated by Congress to be stored within the VistA. This results in one of the richest sources of big data available to surgeons in health services research and an ideal data source for studies needing long-term patient follow-up or questions related to processes of care in surgery. The VHA data include a wide variety of data elements from not only the clinical environment but also from enrollment, benefits, financial, and administrative data all stored within the VHA Corporate Data Warehouse (CDW). Many other sources of information are also available and easily merged into the CDW data by either a social security number or a VHA-specific scrambled social security number (Fig 1).

How Do You Access VHA data?

VHA data are stored as a relational database organized into themed domains. The domains represent specific characteristics of the data pulled from VistA (ie, Pharmacy Outpatient or Vital Signs). To access the data, a researcher must first submit a request through the Data Access Request Tracker (DART), a web-based application. Once the request is reviewed and approved by the National Data System, it is fulfilled through the VA Informatics Networking and Computing Infrastructure (VINCI).

VINCI is also a web-based platform. It provides easy access to the requested data, along with nearly all analytic software that might be needed to manage and analyze the data, all from within the VHA firewall. Access to the software is granted at the same time as data access is granted. VINCI also provides excellent technical support and access to a live VHA data manager, making it the ideal one-stop shop for VHA data analysis.

Available VHA Data Elements for Surgical Research

Given the breadth of data available, VHA data can help address a variety of questions in surgical research. All aspects of a veteran’s health care are captured within the system. These include patient demographics, inpatient stays, outpatient clinic visits, pharmacy orders and fills, laboratory results, vital signs, patient assessments (ie, AUDIT-C), and even aspects of cost. It is important to keep in mind that the VHA data represent an administrative data source, so diagnoses and procedures of interest must be culled from the litany of International Classification of Disease (ICD), version 9 and version 10 codes or Current Procedural Terminology codes found in the inpatient and outpatient domains. However, many resources for accurately coding these data already exist in the literature.2,3

Unique Capabilities of VHA Data

One of the unique capabilities of VHA data is the opportunity for natural language processing (NLP). NLP is available with the assistance of VINCI and can be used in surgical research to alleviate the need for chart abstraction in large data samples. For example, Murff et al4 used NLP to identify several postoperative complications in a sample of 3,000 patients and found greater sensitivity using NLP as compared to the Agency for Healthcare Research and Quality Patient Safety Indicators. More recently, NLP has been used in VHA data to identify ejection fraction values buried within free-text notes5 and even detailed pathology data after a bladder cancer diagnosis.6

VINCI also has the capability of merging in several other data sources, such as Medicare or Medicaid data, which are maintained by the VA Information Resource Center. This allows for a nearly complete view of health care utilization among veterans and is particularly useful for examining the incidence of services in a population. As an example, Hawn et al7 undertook a study, examining the incidence of any noncardiac surgery after cardiac stent implementation. After incorporating Medicare and Medicaid data on the stented population, they found that 29.3% of surgeries in

https://doi.org/10.1016/j.surg.2018.07.029
0039-6060/Publicated by Elsevier Inc.
the following 2 years occurred outside of the VHA. Through administrative data coding, the study team was able to include these non-VHA surgeries in further analyses and found that non-VHA procedures were more often lower complexity elective surgeries as compared with procedures identified in the VHA data. These types of analyses are typically not possible when using other data sources.

Another exciting development in VHA data is the addition of the Million Veteran Program (MVP). The MVP is a national research program examining the effect of genomics on veteran health. This program provides an opportunity to examine precision medicine in the field of surgery. The objective of the MVP is to collect blood samples, a data element not currently captured by VHA data, on one million veterans, roughly 10% of the annual patient population. The MVP data will add more patient assessments and genomic data to an already clinically rich data source. As of 2017, the program has enrolled more than 600,000 veterans and is on target to be completed in 2021.

Summary

VHA data are the ideal big data answer for surgical studies needing long-term follow-up or addressing questions related to processes of care in surgery. Given the large sample size, it is also a great opportunity to study rare outcomes and rare exposures or for studies requiring more than just clinical data elements. The VHA data does come with the limitation of generalizability and a need for administrative data coding, but its robustness far outweighs these limitations. With options for NLP, the ability to easily merge in other data sources, and the collection of genomics data from the MVP, VHA data are the perfect mix of volume and variety for surgical health services research studies.
Laura A. Graham, PhD, MPH
Center for Innovation to Implementation (Ci2i), VA Palo Alto Health Care System, CA
Stanford-Surgery Policy Improvement Research & Education (S-SPIRE), Stanford University, CA

Reprint requests: Laura A. Graham, PhD, MPH, Center for Implementation to Innovation, VA Palo Alto Healthcare System, 795 Willow Road, Bldg 324, Rm B-129, Menlo Park, CA 94025.
E-mail address: lagraham@stanford.edu.

Accepted 25 July 2018

References