Advancing veterans’ healthcare using electronic data: Lessons learned from researchers in the field

The Department of Veterans Affairs (VA) provides nationwide access, lifetime coverage, and an integrated care structure to its enrollees. Those key aspects of VA healthcare—together with data contained in VA’s electronic information systems supporting over eight million veterans—provide unique opportunities to study processes, outcomes, and costs of care. Recently, for example, VA data have been used to study outcomes associated with acute postoperative inpatient rehabilitation and care in specialized rehabilitation bed units after lower-limb amputation [1–2], medication adherence and relapse among patients discharged from a VA posttraumatic stress disorder (PTSD) treatment program [3], the provision and costs of assistive technology devices to veterans after stroke [4], and use of mental health services by veterans disabled by auditory disorders [5].

UNDERPINNINGS OF SINGLE-TOPIC ISSUE ON DATA QUALITY

In 1998, the VA Health Services Research & Development Service funded the VA Information Resource Center (VIReC) to facilitate the use of VA administrative data for research. Since that time, VIReC has developed and disseminated research-relevant information about VA databases and information systems. Knowledge-building activities include data quality investigations conducted within VIReC. This single-topic issue recognizes the vast stores of information about VA data that VA investigators generate nationwide, in the normal course of their data analysis activities, and seeks to leverage that information to advance electronic-data-based research.

A call for abstracts was issued in the spring of 2009 for investigations of the quality and research utility of electronic data used in research to advance the care of veterans. VA investigators responded—many more worthy abstracts were submitted than could be included in one JRRD issue. Reflecting strong research-community support for VA research, 54 scholars readily accepted invitations to peer review and almost 100 individuals participated as authors, reviewers, or editors.

VA researchers have long been major contributors to the methodological literature in the area of administrative data use [6–17]. This issue advances that literature by focusing specifically on the quality and research utility of databases used in veterans research. We view data quality—roughly, the completeness and accuracy of the data collected and entered—as a necessary but not sufficient feature of a “research-useful” data set [18]. The research
utility of a data set entails additional factors such as coverage of the population of interest and degree of validity achievable when specific data elements are used as proxies for important but absent information. Research utility questions, therefore, relate to the inevitable challenges involved in the use of secondary data—i.e., data that were collected for a different purpose—rather than a deficiency in the data per se.

**VITAL VETERANS RESEARCH AND VA HEALTH INFORMATION SYSTEMS**

The questions of data quality and research utility addressed in the included articles were encountered during research tackling key issues in the care and rehabilitation of wounded war veterans. Promoting high-quality care and providing the tools for independent living to veterans with spinal cord injury (SCI) [19], polytrauma [20], PTSD, substance-use disorder and other mental health conditions [21–23] and to others needing rehabilitative and assistive services [24–25] are high VA priorities. Investigating obesity and chronic kidney disease and its antecedents are important because these are among the most common chronic conditions afflicting veterans seeking care in the VA [26–27]. Understanding the healthcare needs and ensuring accessible care for women and other vulnerable populations [23,28–30] are among the VA’s top research priorities. Identifying cost-effective approaches to veterans’ healthcare [25,29,31] is critical to long-term viability of the VA healthcare system. Studies included in this issue were derived from research investigating these issues using data from sources inside and outside the VA.

The included articles also spotlight the great value that VA’s health information systems have for research. Chief among the data sources mined for research is the electronic medical record (EMR), portions of which are extracted to create the VA National Patient Care Database, providing clinical and health services utilization information. Condition-based registries maintained for clinical and administrative purposes and other case-reporting systems provide additional clinical and treatment detail and so are also rich data sources. Data originating outside the VA are critical for learning about care received and costs of care for veterans obtaining services in the private sector and help to fill a gap in information for many studies. In addition to the National Patient Care Database [19–21,24,30], data sources evaluated include the Veterans Health Administration Corporate Data Warehouse [26], the VA’s National Spinal Cord Dysfunction Registry [19], the Minimum Data Set (long-term care resident assessment) [24], Medicare’s End-Stage Renal Disease Medical Evidence Report [27], Medicaid and Medicare claims [29–30], and the Medicare Current Beneficiary Survey [31].

**SPOTLIGHT ON DATA QUALITY AND RESEARCH UTILITY STUDIES**

Several studies examined the completeness and/or validity of these data resources. Some common observations emerge from among this set of studies. First, findings suggest substantial variability in record completeness across VA facilities and regional networks [20,22]. Both the inconsistency itself and the inferior quality found in some locations have clear implications for patient care quality as well as validity of performance measurement and research conclusions. Because all locations are within the same healthcare system, the variability is unlikely to be attributable to heterogeneity in mission, goals, economic incentives, or data-entry system. Although the reasons for the inconsistency in quality are unclear and probably multifactorial, the recognition of variability nonetheless presents opportunities to improve systemwide performance.

Second, data collected for purposes peripheral to direct patient care or billing seem to be particularly prone to inaccuracy [19–20,24,27,30]. Several studies examined data collection systems that require manual data entry by clinical staff, separate from routine medical record documentation [19,24,27]. Not surprisingly, accurate and complete data entry will receive lower priority when not essential for patient care. In other cases, lack of an administrative
mandate for data collection, which otherwise may provide incentives and technical controls, may contribute to incomplete data.

Unlike data quality, the research utility of a data set depends somewhat on the skill and experience of the researcher. A thorough understanding of the data and time and effort spent constructing and evaluating the validity of operational definitions, for example, can substantially alter measurement of important variables. B. Smith et al. exemplify this truth as they share results of extensive experience in employing VA electronic data to identify and characterize veterans with SCI [19]. The point is also illustrated by Frayne et al., who tested a number of algorithms to construct a length-of-stay measure and found that estimates of mental illness-related disparity depended on the choice of algorithm [23]. Several articles investigated methods of case identification using administrative data [19–22,24]. Carlson et al. examined the feasibility of using external cause-of-injury codes to identify subsets of the injured patient population [20]. Harris et al. determined that use of VA clinic stop codes alone is inadequate to identify patients receiving outpatient treatment for substance use disorder but that adding diagnosis codes to the case definition greatly improved ascertainment [22]. Berlowitz et al. evaluated whether administrative data could provide more accurate information than the Minimum Data Set about the presence of comorbid conditions among residents in long-term and rehabilitation care settings [24]. Washington et al. reported their experience in combining data from several sources to create a sampling frame of women veterans [28], and Bangerter et al. investigated the accuracy of an algorithm to identify patients with PTSD in administrative data [21]. M. Smith et al. described techniques used to link patient-level information in the EMR and the National Prosthetics Patient Database (NPPD) and concluded that use of the NPPD can improve assessment of resource utilization, but they caution researchers against use of the NPPD to identify the universe of patients receiving specific prosthetics services [25]. Hendricks et al. explored the value of including Medicaid data in veterans studies and shared insights gained from extensive experience in use of VA and Medicaid data for research [29]. All these studies contradict a common impression that use of secondary data for research is straightforward and uncomplicated.

As the VA continues to respond to the healthcare needs of U.S. veterans and, in particular, to the unique and evolving needs of those returning from Afghanistan and Iraq, research that identifies those needs, evaluates the VA’s response, and assesses patient outcomes is essential. VA electronic data will continue to be a vital tool for that research and analysis. The broad goal of the studies reported in this issue is uniform: to achieve excellence in the conduct of current and future research using VA data. We hope and anticipate that the results presented in this single-topic issue, through an immediate positive impact on research going forward, will advance VA healthcare toward the ultimate goal of each veteran’s optimum health and well-being.

Elizabeth Tarlov, RN, PhD;* Kevin Stroupe, PhD
Edward Hines Jr. VA Hospital—VA Information Resource Center, Hines, IL

*Email: elizabeth.tarlov@va.gov

REFERENCES


22. Harris AHS, Reeder RN, Ellerbe L, Bowe T. Are VHA administrative location codes valid indicators
This article and any supplementary material should be cited as follows:

DOI:10.1682/JRRD.2010.08.0148