
Original Articles

Measurement in Veterans Affairs Health Services Research: Veterans as a Special Population

Robert O. Morgan, Cayla R. Teal, Siddharta G. Reddy, Marvella E. Ford, and Carol M. Ashton

Objective. To introduce this supplemental issue on measurement within health services research by using the population of U.S. veterans as an illustrative example of population and system influences on measurement quality.

Principal Findings. Measurement quality may be affected by differences in demographic characteristics, illness burden, psychological health, cultural identity, or health care setting. The U.S. veteran population and the VA health system represent a microcosm in which a broad range of measurement issues can be assessed.

Conclusions. Measurement is the foundation on which health decisions are made. Poor measurement quality can affect both the quality of health care decisions and decisions about health care policy. The accompanying articles in this issue highlight a subset of measurement issues that have applicability to the broad community of health services research. It is our hope that they stimulate a broad discussion of the measurement challenges posed by conducting “state-of-the-art” health services research.

Key Words. Measurement, special populations, veterans

The U.S. Department of Veterans Affairs Health Services Research and Development (VA HSR&D) service has been a leader in focusing on system-wide measurement excellence, in part because veterans represent a distinct and special population. Veterans are a large and highly selected group with an increasingly recognized cultural identity. Further, eligible veterans are served by a large, nationally integrated health care system. Thus, the veteran population and the VA health system embody a microcosm in which measurement quality must be assessed and addressed. In this article, we briefly discuss the consequences of poor measurement and the importance of assessing measurement quality across the varied populations in which health services

research is conducted, using the veteran population as a case example. We then introduce a series of related articles, each of which focuses on a measurement issue important in current health services research.

MEASUREMENT IN HEALTH SERVICES RESEARCH

Data that are unreliable or have poor validity can lead to erroneous and nongeneralizable study results through a combination of low statistical power and lack of sensitivity in data analyses, biases in statistical conclusions, and biases in estimates of prevalence and risk (Skinner, Teresi, and Holmes 2001). These errors can affect our understanding of therapeutic effectiveness by restricting our ability to detect an intervention's effect, and distort our assessments of the epidemiology of medical conditions by biasing our assessment of different subpopulations of patients.

It is widely recognized that measurement properties such as reliability and validity are both sample- and purpose-dependent (Anastasi 1998). That is, they vary across the populations and settings in which measures are used. Typically, researchers are most familiar with these issues in the context of measurement with self-report instruments, surveys, or scales. On scales, for example, individual items may differ across populations in terms of how they relate to the underlying constructs being measured, and the constructs themselves may shift across populations. Measures may be affected by differences in demographic characteristics (e.g., age, socioeconomic status, location), illness burden, psychological health, or cultural identity. Consequently, a scale developed to assess communication ability in Anglo Americans may not be as effective when used with African Americans or Hispanic Americans; a scale may not work as well with individuals raised in a rural setting as with those raised in an urban one; or the properties of a scale developed in a sample of young female patients may not generalize when the scale is used with older males. Similarly, the measurement properties of scales may vary according to

Address correspondence to Robert O. Morgan, Ph.D., Houston Center for Quality of Care and Utilization Studies (152), Veterans Affairs Medical Center, 2002 Holcombe Blvd., Houston, TX 77030. Robert O. Morgan, Cayla R. Teal, Ph.D., and Carol M. Ashton, M.D., M.P.H., are with the VA HSR & D, Houston Center for Quality Care and Utilization Studies, Houston, TX, and with Baylor College of Medicine, Houston, TX. Siddharta G. Reddy, M.P.H. is with the VA HSR & D, Measurement Excellence and Training Resource Information Center (METRIC), Houston, TX. Marvella E. Ford, Ph.D., Associate Director, is with the Health Disparities Research Program, Medical University of South Carolina, Charleston, SC.

how they are used. For example, a measure developed for assessing cross-sectional group differences in health status may be inadequate as an instrument for measuring change over time for a particular individual. When measurement is conducted via survey methodology, these vulnerabilities may be compounded by biased nonresponse to the survey or partial completion of survey items.

The need to verify measurement properties extends beyond “traditional” psychometric applications (e.g., reliability or validity of survey or other self-report measures) and beyond the characteristics of the population we are attempting to study. For the U.S. population in general, there are substantial differences among the health care systems in which individuals seek care. These differences may affect entry into the system (e.g., access), therapeutic decisions (e.g., quality), and availability of end-points (e.g., outcomes). Thus, as health services researchers, measurement and our resulting research findings are influenced by features of the health care system. Health services research incorporates measurements obtained through direct observation, self-report, or from administrative or medical records (e.g., illness classification, health care use, morbidity, mortality). Attention to measurement quality necessarily includes design issues (e.g., formatting and administration of measurement instruments), settings in which measurement is conducted (e.g., at a physician’s office versus a hospital setting, or at home), and the source from which the measures are obtained (e.g., self-report by an individual, observer rating, administrative or medical record).

VETERANS AS A SPECIAL POPULATION

Research with veterans and within the VA health care system serves as a case example of how measurement can be affected by the issues raised above. The population characteristics of veterans reflect the characteristics of the armed forces in which they served. As a group they are predominantly male, and more educated and better off financially than the general U.S. male population (Klein 2001; Klein and Stockford 2001). The male veteran population is projected to decline substantially (approximately 27 percent) between 2005 and 2020. In contrast, the female veteran population is projected to increase by 12 percent over the same period, reflecting the changing gender composition of the military (U.S. Department of Veterans Affairs 2000). How these changes will be altered by the conflict in Iraq is unclear. Like the Gulf War, the war in Iraq has seen substantial mobilization of military reserve units; and although

training for the Reserves or National Guard unit does not entitle an individual to veterans' benefits, activation for service does.

Currently, the majority of veterans belong to the age cohorts who served in World War II, the Korean War, and the Vietnam War. The median age of all veterans is 55, with veterans comprising a majority of all civilian males older the age of 65 (U.S. Department of Veterans Affairs 2000). However, this proportion varies by race, with veterans accounting for over 60 percent of white males older the age of 65, but only 45 and 35 percent of African-American and Hispanic males in the same age range (Bureau of the Census 2001).

VA Health System Users

Veterans who use VA Health Administration (VHA) services are an even more highly selected population of veterans. Although all honorably discharged veterans are eligible to receive care through VHA facilities, priority for care goes to veterans who have service-related disabilities, who are in certain veteran groups (e.g., prisoners of war), or who meet specific criteria for financial need. Other qualified veterans (i.e., honorably discharged veterans) may also be able to receive services at VA health care facilities, albeit with a lower priority and potentially with additional or higher copayments. Possibly as a consequence of VA system priorities, VA users appear to be poorer, older, less well educated, more likely to be unemployed or underemployed, more likely to be African American, and more likely to report poorer physical and mental health and more chronic health conditions than either the general population or veterans who do not use the VA health care system (Kazis, Miller, and Clark 1998; Agha, Lofgren, and Van Ruiswyk 2000).

Use of the VA health system also appears to be influenced by an individual's self-identity as a veteran. For example, preference for VA outpatient care, as opposed to non-VA care, appears significantly associated with combat exposure, war era (e.g., WWII, Korean War), rating of military experience, membership in veterans' organizations, and veteran influence in daily life (Harada, Damron-Rodriguez, and Villa 2002).

Thus, the sociodemographic differences between veterans who do and do not use the VA health system raise inevitable measurement questions. For example, are the measurement properties of questions about health and health care use the same in veterans who do and do not use the VA health system? Do individuals who use multiple systems of care, e.g., the VA system and Medicare, respond to questions about access to care the same way that individuals using a single system do? If VA users tend to more strongly identify with their

veteran status than veterans who do not use the VA, does this stronger “veteran identity” mute other possible influences on measurement properties, for example racial, ethnic, and gender influences?

The VA as a Specific Health Care Setting

The process of care within the VA system is shaped not only by the characteristics of the veterans who seek care within it, but also by the organizational structure and policies specific to the VA. The VA is distinguishable from federally funded health insurance programs, such as Medicare and Medicaid, in that the VA provides medical care directly, rather than financing medical care provided through the private sector. VA medical care facilities are not reimbursed for specific episodes of care. Thus, administrative records describing care have a very different purpose than comparable data obtained from other health care systems. For example, International Classification of Diseases (ICD-9) and Current Procedural Terminology (CPT) codes available through the Medicare claims files are used to justify reimbursements to providers under the traditional Medicare fee-for-service (FFS) system. Consequently, there is financial pressure on provider staff to not “under-code” episodes of medical care. Historically, similar financial pressures have not existed within the VA system, at least to the same extent. This highlights a general point: validation studies of medical coding that are carried out within one system, e.g., Medicare, do not necessarily translate to coding performed within another, such as the VA.

The VA is designed as an integrated national system of care. Thus, policies and guidelines are developed with the expectation that they will be applied throughout the national VA system of medical centers and outpatient clinics. In this respect, the VA system resembles a very large, staff model managed care organization. Thus, while there is variability in how well guidelines are followed at any local VA installation, there is a specific, system-wide effort toward standardization. With the exception of periodic auditing of Medicare providers to monitor compliance to coding regulations, the same cannot be said of the Medicare system. This means that within the Medicare system, in addition to coding differences resulting from differences in pressures on the coders themselves, there may also be system-level differences in variability across facilities. Thus, studies using administrative files to examine the epidemiology of medical conditions, or which use administrative files for case-finding, need to pay specific attention to the system processes that affect how individual records of care are coded.

The influence of system characteristics extends to other types of measures as well. Until relatively recently, the VA has emphasized the development of facilities that integrate inpatient and ambulatory care services into the same physical structure, i.e., VA medical centers, and with a common source of financing. This is a profoundly different organization of care than exists in the private sector. Leaving aside any discussion of the relative merits of each approach to care, the different organizations of care clearly may have an effect on the meaning of frequently used measures of care quality. For example, if VA physicians use different criteria for deciding when to admit a patient, do admission rates within the Medicare and VA systems mean the same thing?

Thus, the population of veterans represents a distinct group of individuals, many of whom seek care through a health care system especially designed for their needs. As such, they serve as an exemplar of a special population, and provide both a need and an opportunity for examining population and setting specific influences on the measurement process within health services research.

THE VA “METRIC”

The VA HSR&D service has recognized that providing researchers with basic measurement information and tools may have profound effects on improving the quality of measurement in research. Thus, in 2001 the VA HSR&D Service funded the Measurement Excellence Initiative (MEI). The aim of this initiative was to gather the expertise of psychometricians, research scientists, and students in the related sciences, to provide a web-based measurement resource to the VA research communities. In July 2003, the VA HSR&D service expanded the scope of the MEI and designated it as a resource center, named the Measurement Excellence and Training Resource Information Center, or METRIC (<http://www.measurementexperts.org/>).

METRIC serves VA researchers by assisting researchers to more accurately measure the health, social, and economic condition of the veterans who use the services of the VHA. Increasingly sound data, in turn, enables VA organizations to improve the quality of care for veterans by making evidence-based decisions in the form of clinical-, organizational-, or system-level modifications. Thus, the mission of METRIC is to disseminate information to health services researchers regarding all phases of the measurement process. This includes assisting researchers in finding and evaluating measurement instruments; providing education regarding how to interpret and use

measurement information and how the quality of their measurements are influenced by their study design, study setting, measurement methods, and source of their data; facilitating the sharing of measurement knowledge across the VA research and development community, particularly with regard to the integration of newer measurement approaches (e.g., item response theory and computer adaptive testing [CAT]); and ultimately, advancing measurement science through research.

The articles in this supplement, described below, reflect the diversity of measurement interests held by METRIC-affiliated researchers.

ARTICLES IN THIS SUPPLEMENT

On March 14, 2002, METRIC researchers convened 17 experts in Houston to debate some difficult measurement topics. The expert panel consisted of seven national measurement experts (Drs. Lee Sechrest, Jack Clark, Robert DeVellis, Jay Magaziner, Colleen McHorney, Evelyn Perloff, and Stephen Zyzanski). In addition the panel included 11 METRIC measurement experts (Drs. Carol Ashton, Karon Cook, Marvella Ford, P. Adam Kelly, Robert Morgan, Kimberly O'Malley, A. Lynn Snow, Paul Swank, Nelda Wray, and Mary York).

The measurement topics covered at that meeting were ones we had found to present significant problems for clinicians and researchers in their efforts to improve public health. The ensuing discussions by the expert panel led to the development of the articles included in this supplement. We believe that measurement is the foundation on which health decisions are made, both inside and outside the VA health care system, yet clinicians and researchers are frequently not given training that enables them to address difficult measurement issues. The articles in this supplement are relevant not only for measurement issues involving the VA system and the veteran population, but also pertain to health services research in general. These issues fall into three major categories:

- (1) Measurement validity within a specific context or setting;
- (2) Identifying populations of interest and appreciating measurement challenges associated with these populations; and
- (3) Consideration of alternative methodologies to measure the same construct.

Measurement Validity within a Specific Context or Setting

The first three articles discuss the importance of, and barriers to, ensuring the validity of findings. The goal of the first manuscript, "Validity of Measures Is

No Simple Matter,” by Lee Sechrest, is to promote a better understanding of the nature of measurement, the special problems posed by measurement in the social sciences, and the inevitable limitations on inferences in science (so that results of any sort are not overinterpreted). Sechrest proposes that judgments about degree of measurement validity can be guided by careful consideration of the measurement process, measurement context, and purpose of the measurement. The manuscript highlights the difficulty of measurement using blood pressure as an example. Suggestions for the promotion of good measurement practice are offered.

The second manuscript, “Integrating Validity Theory with Use of Measurement Instruments in Clinical Settings,” by Kelly, O’Malley, Kallen, and Ford, challenges the rationales investigators often give for their decisions to use specific measurement instruments in clinical settings. The authors offer a step-by-step, three-level decision rubric that focuses on key considerations in assessing validity. At each level, specific questions are posed and solutions grounded in validity theory are suggested. This manuscript will assist investigators in clinical settings organize and focus evaluation of, and justification for, using specific measurement instruments.

ICD codes are now used in many applications, including research, reimbursement, identification of medical errors, and denoting causes of death on death certificates. Because ICD-9-CM and ICD-10-CM codes are widely applied to “measure” diagnoses, their accuracy is of considerable interest. Understanding sources of error in the application of these codes is critical to the evaluation of their usefulness and limitations. In the third article, “Measuring Diagnoses: ICD Code Accuracy,” O’Malley, Cook, Price, Raiford Wildes, Hurdle, and Ashton conceptualize diagnostic coding as a measurement problem. The authors summarize the process for assigning diagnostic codes, identify sources of errors in the process, summarize research related to code accuracy, and review methods for quantifying the accuracy of diagnostic codes. By understanding potential sources of errors in the process of assigning diagnostic codes, code users (i.e., researchers, clinicians, payers, etc.) can make better decisions about the usefulness of the codes in various applications.

Identifying Populations of Interest and Appreciating Measurement Challenges Associated with these Populations

The next two articles address issues involved in identifying populations of interest. The fourth manuscript, “Measurement Issues in Health Disparities Research,” by Ramírez, Ford, Stewart, and Teresi provides an overview of measurement instruments in diverse populations. Members of minority

groups have higher rates of morbidity and mortality than their nonminority counterparts for almost all categories of disease. Racial and ethnic health disparities may be because of comorbidity, access to care, attitudes and perceptions, or disease etiology. However, in order to conduct comparative studies in a meaningful manner, investigators must face the challenge of addressing cross-cultural measurement equivalence, including issues of validity and reliability.

In turn, valid and reliable measures of race and ethnicity are needed to ensure accurate assessment of disease prevalence and incidence, and accounts of appropriate health services utilization, in different population groups. Thus, the purpose of the fifth manuscript, "Conceptualizing and Categorizing Race and Ethnicity in Health Services Research," by Ford and Kelly, is threefold. First, this manuscript provides an overview of different methods currently employed to assess the constructs of race and ethnicity. Second, the authors illustrate consistent standards for measuring these constructs. The authors conclude with suggestions for measurement methods that would improve future research.

Consideration of Alternative Methodologies to Measure the Same Construct

The final set of articles examines the properties and use of two important measurement methodologies. In "Proxies and Other External Raters: Methodological Considerations," Snow, Cook, Lin, Morgan, and Magaziner review past research on proxy reports and examine ways to increase the reliability and validity of these types of reports. Many of the constructs health services researchers attempt to measure are those that have no single objective gold standard (e.g., quality of life, pain). Thus, the accrual of validity evidence for a health services research assessment instrument is often a daunting task. The difficulty of this task is increased further when it is not possible to obtain a self-report from the patient, or when the self-report is suspect. Thus, the use of reports from external raters (e.g., family members, clinicians) is a necessary strategy for establishing evidence of validity. The authors differentiate between external rater reports gathered for the purpose of *substituting* for self-report and those gathered to *supplement* the self-report. The authors explain why appropriate use of externally rated data requires careful consideration of the nature of the data and how it will be analyzed and interpreted.

In the seventh and final article, "Dynamic Assessment of Health Outcomes: Time to Let the CAT Out of the Bag?" Cook, O'Malley, and Roddey discuss CAT in the assessment of patient reported outcomes. CAT-based

measures are “dynamic” in that the set of items presented to a patient is individualized based on continually updated estimations of the patient’s level of the trait being measured. This tailored approach to assessment returns increased measurement precision without increased response burden. The article by Cook, O’Malley, and Roddey describes what CAT is (and is not) and discusses its promise and its problems.

SUMMARY

Measurement is the foundation on which health decisions are made, both inside and outside of the VA system. Poor measurement quality can affect the quality of health care decisions, contributing to over- or undertreatment of medical conditions, and, at a system level, can affect decisions about health care policy. While there are numerous measurement issues that deserve consideration (e.g., Rasch approaches to measurement, the linkage or comparison of data from different administrative databases), the accompanying articles highlight a subset of measurement issues that have applicability to the broad community of health services research. It is our hope that they stimulate a broad discussion of the measurement challenges posed by conducting “state-of-the-art” health services research.

ACKNOWLEDGMENT

This material is based upon work supported by the Health Services Research and Development (HSR&D) Service of the Office of Research and Development, Department of Veterans Affairs.

REFERENCES

- Agha, Z., R. P. Lofgren, J. V. VanRuiswyk, and P. M. Layde. 2000. “Are Patients at Veterans Affairs Medical Centers Sicker? A Comparative Analysis of Health Status and Medical Resource Use.” *Archives of Internal Medicine* 160 (21): 3252–7.
- Anastasi, A. 1998. *Psychological Testing*, 6th ed. New York: Macmillan Publishing Company.
- Bureau of the Census 2001. *Census 2000*. Washington, DC: U.S. Department of Commerce.
- Harada, N. D., J. Damron-Rodriguez, V. M. Villa, D. L. Washington, S. Dhanani, H. Shon, M. Chattopashyay, H. Fishbein, M. Lee, T. Makinodan, and R. Andersen.

2002. "Veteran Identity and Race/Ethnicity: Influences on VA Outpatient Care Utilization." *Medical Care* 40 (suppl, 1): I117-28.
- Kazis, L. E., D. R. Miller, J. Clark, K. Skinner, A. Lee, W. Rogers, A. Spiro III, S. Payne, G. Fincke, A. Selim, and A. Linzer. 1998. "Health-Related Quality of Life in Patients Served by the Department of Veterans Affairs: Results from the Veterans Health Study." *Archives of Internal Medicine* 158 (6): 626-32.
- Klein, R. E. 2001. *The Changing Veteran Population, 1990-2020*. Washington, DC: U.S. Department of Veterans Affairs.
- Klein, R. E., and D. D. Stockford. 2001. *Data on the Socioeconomic Status of Veterans and on VA Program Usage*. Washington, DC: U.S. Department of Veterans Affairs.
- Skinner, J. H., J. A. Teresi, D. Holmes, S. M. Stahl, and A. L. Stewart. 2001. "Measurement in Older Ethnically Diverse Populations: Overview of the Volume." *Journal of Mental Health and Aging* 7 (1): 5-8.
- U.S. Department of Veterans Affairs 2000. *VetPop2000*. Washington, DC: U.S. Department of Veterans Affairs.