

Data Resources in the Department of Veterans Affairs

CHARLES MAYNARD, PHD
MICHAEL K. CHAPKO, PHD

OBJECTIVE — To provide an overview of databases that are maintained by the Department of Veterans Affairs (VA) and are of relevance to investigators involved in epidemiologic, clinical, and health services research.

RESEARCH DESIGN AND METHODS — We reviewed both national and local VA databases and identified their strengths and limitations. We also referenced specific studies that have assessed the validity and reliability of VA databases.

RESULTS — There are numerous national databases housed at the Austin Automation Center in Austin, Texas. These include the Patient Treatment File (hospital abstracts), the Outpatient Care File, the Beneficiary Identification Record Locator System death file for assessing vital status, and the Decision Support System, which provides integrated clinical and financial information for managerial decision making. The major limitation of these databases is that clinical detail below the level of ICD-9-CM diagnosis and procedure codes is not uniformly available nationally. These databases offer an excellent opportunity to monitor the health of veterans over time because they track all inpatient and outpatient utilization in the VA. However, at the local or medical center level, the Veterans Health Information and Systems and Technology Architecture contains extensive clinical information, but has fewer patients, varies in format across medical centers, and poses difficulties with data extraction for statistical analysis.

CONCLUSIONS — Both local and national VA databases are valuable resources for investigators who have interests in a wide array of research topics, including diabetes. The potential for investigating important scientific questions with VA databases becomes greater as communications and database management technologies improve.

Diabetes Care 27 (Suppl. 2):B22–B26, 2004

The Department of Veterans Affairs (VA) has a vast array of databases at both the national and local levels. These resources are of potential value to investigators who do epidemiologic, clinical, and health services research. The purpose of this article is to review these resources and to assess their relevance for those interested in the epidemiology of

diabetes. This review will focus primarily on the strengths and limitations of the national databases, including the Patient Treatment File (PTF), the Outpatient Care Files (OPC), and the Beneficiary Identification Record Locator System (BIRLS) death file. Less attention will be given to local files, including the Veterans Health Information Systems and Tech-

nology Architecture (VistA) and the Consumer Health Information and Performance Set (CHIPS) database in the Veterans Integrated Service Network (VISN) or region 20. Finally, recent developments in the national databases will be described.

VA national databases

Hospital discharge data has been collected from VA medical centers since 1970 and clinic encounters since 1980. In addition to maintaining data at the local level, databases of national relevance are housed at the Austin Automation Center in Austin, Texas (Fig. 1). These include the PTF and OPC files, as well as the Decision Support System (DSS), which provides integrated clinical and financial information for managerial decision making. For epidemiologists, clinical, and health services researchers, the PTF and OPC files are of greatest interest. Collectively, the PTF and OPC are subsets of the National Patient Care Database. Detailed descriptions of these files and their contents are available on the VA Information Resource Center (VIREC) website (www.virec.research.med.va.gov) (1). VIREC provides information on a variety of topics, including the mechanics of connecting to Austin, programming tips, and an inventory of databases of interest to researchers. Recently published studies have used existing VA databases to describe hospital and clinic utilization, as well as geographic differences in the use of VA hospitals and clinics (2,3). In addition, an article discussing the importance of VA databases for epidemiologists has been published (4).

Hospital data. The PTF files are analogous to hospital discharge abstracts and include basic demographic data as well as principal, primary, and nine secondary ICD-9-CM codes. In the VA, the primary diagnosis code refers to the condition that accounted for the majority of the hospital stay. The principal diagnosis code, which was added in FY1997, refers to the diagnosis that led to hospitalization. In addition to the main PTF file, surgeries and procedures files are also available. The surgeries file contains ICD-9-CM proce-

From the HSR&D Center of Excellence, VA Puget Sound Health Care System, and the Department of Health Services, University of Washington, Seattle, Washington.

Address correspondence and reprint requests to Charles Maynard, PhD, VA Puget Sound Health Care System (152), 1660 S. Columbian Way, Seattle, WA 98108. E-mail: cmaynard@u.washington.edu.

Received for publication 1 July 2003 and accepted 25 July 2003.

Funding for this supplement was provided by The Seattle Epidemiologic Research and Information Center and the VA Cooperative Studies Program.

Abbreviations: BIRLS, Beneficiary Identification Record Locator System; CHIPS, Consumer Health Information and Performance Set; CPRS, Computerized Patient Record System; DSS, Decision Support System; OPC, Outpatient Care Files; PTF, Patient Treatment File; VA, Department of Veterans Affairs; VIREC, VA Information Resource Center; VISN, Veterans Integrated Service Network; VistA, Veterans Health Information Systems and Technology Architecture.

A table elsewhere in this issue shows conventional and Système International (SI) units and conversion factors for many substances.

© 2004 by the American Diabetes Association.

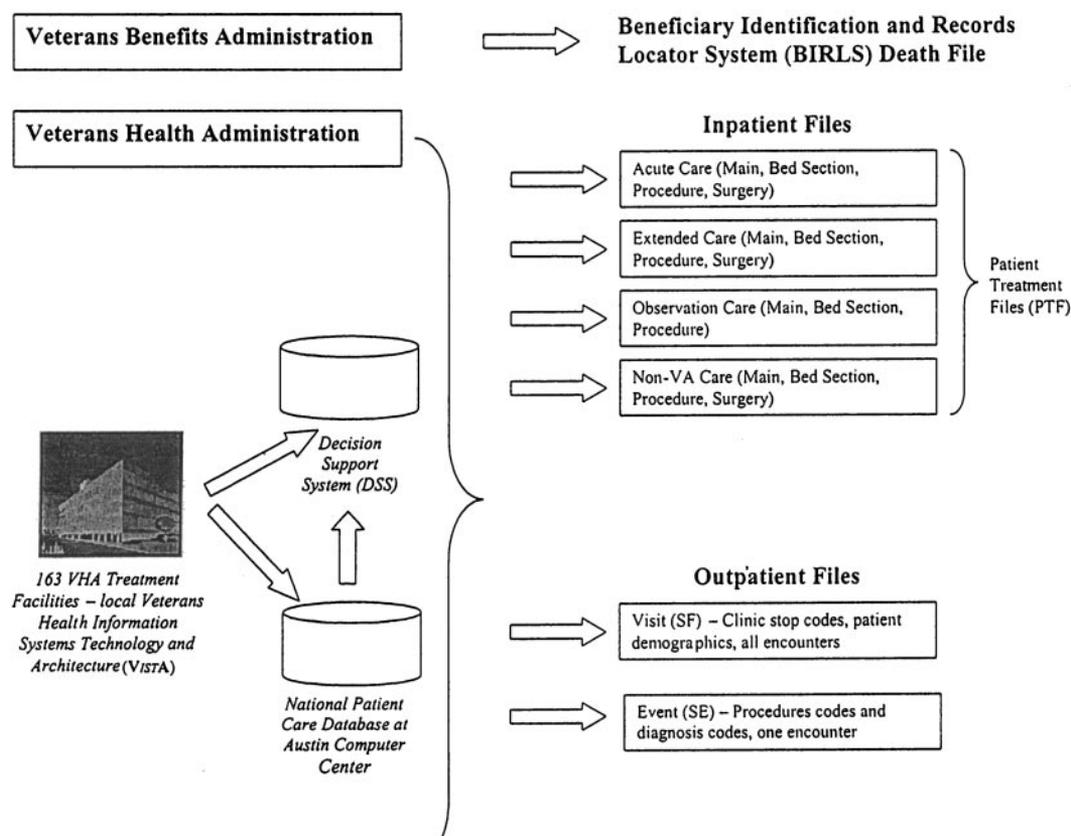


Figure 1—Data flow from Veterans Health Administration Medical Centers to the Austin Automation Center: inpatient data (adapted from Inpatient Data Resource Guide) and outpatient data (adapted from Outpatient Data Resource Guide), FY2000

cedure codes for surgical procedures (e.g., coronary artery bypass surgery) performed in the operating room, whereas the procedures file includes codes (e.g., cardiac catheterization) for procedures performed outside the operating room. The surgeries file has an observation for each surgery performed during an episode of care, and up to five codes per surgery may be listed. The procedures file includes an observation for each day procedures occurred during the hospitalization, and up to five procedures may be listed. There is a maximum of 32 observations per hospitalization or episode of care resulting in a maximum of 160 procedure codes. In addition to hospital or acute care, the PTF includes extended care (VA nursing home), observation (outpatient surgery), and non-VA care (care provided in facilities with which the VA contracts).

Starting in 1970, the main PTF, surgeries, and procedures files are available for each fiscal year (October through September). While the contents of the files have changed from year to year, defini-

tions of most items remain constant over time. Data are stored in the Statistical Analysis System (SAS) and can be downloaded from the Austin mainframe to a PC in SAS export files for data analysis. With appropriate file conversion software, these export files can be converted to a wide variety of statistical software packages.

Outpatient data. Three types of OPC files contain information on all patients who are seen in VA outpatient clinics. Since 1980, the visits file has provided demographic information, as well as clinics visited on a given day. These files are arranged by visit day with one visit per patient per 24-h day. Since 1986, all visits are available. Added in 1990, the ambulatory procedures file contains CPT codes for procedures performed during a given visit. A diagnosis file was added in 1997 and contains one primary and nine secondary diagnosis codes. These files contain over 30 million records per year. In 2000, an events file combining information from the outpatient procedures and diagnosis files was implemented. Figure 1

shows the flow of data from the local medical center to the national OPC File.

Vital status data. Of particular interest to epidemiologists is the BIRLS death file, which contains current vital status information on veterans who receive VA benefits (5). The larger BIRLS file is used primarily to collect information about veterans who have applied for VA benefits, veterans discharged from military service since March 1973, Medal of Honor recipients, and service members with accounts for VA education benefits. From this larger file, the death file is created. The file is updated quarterly from information gathered from a variety of information sources inside and outside the VA, including the Social Security Administration. This information ensures benefits are distributed only to living veterans or their eligible dependents. The file is in SAS format, can be linked with the PTF and OPC files, and includes date of death, but has no information on the cause or circumstances of death. Veterans who die, but who are not included in the larger BIRLS file, will not appear in the BIRLS

death file. This is estimated to be ~5–10% of all veterans.

Limitations. A major limitation of these VA databases is that they lack important clinical detail that cannot be determined from ICD-9-CM codes. For example, in the case of diabetes, it is not always possible to distinguish type 1 from type 2 with diagnostic codes. For patients undergoing procedures such as coronary artery bypass surgery, it is often difficult to distinguish complications of a procedure from conditions existing before the procedure. Congestive heart failure and acute myocardial infarction are two examples of this ambiguity. Moreover, administrative databases generally do not have laboratory or pharmacy data that help to define whether an individual has been or is being treated for a given condition such as diabetes, although this information is available at local facilities. However, the VA does have the national Pharmacy Benefits Management database, which includes pharmacy and laboratory data that can be linked to data obtained from the PTF and OPC files. Information about this program is available from the VIREC website.

In addition to these general limitations, there are other shortcomings specific to VA databases. Not all hospital and outpatient care provided to veterans is captured in VA databases. Many veterans receive care in the private sector as well as from the VA. This can be remedied in part by merging databases from Medicare and the VA to determine where veterans 65 years and older receive their care (6,7). Second, despite being a valuable resource, the BIRLS death file does not contain information on cause of death. For investigators interested in these data, it is necessary to either use the National Death Index or obtain death certificates from the appropriate governmental agencies, usually state health departments. Several studies have shown excellent agreement between BIRLS death records and vital status from Social Security Administration and Washington State death records (5,6,8,9). A third limitation relates to technical matters. Since outpatient utilization is contained in two or three very large files for each year, working with these data can be difficult because merging of large files is often required.

A recent report on the reliability of VA databases has revealed other weaknesses (10). Reliability was adequate for demographics and selected diagnoses, but was

inadequate for identifying clinic type. When compared with medical charts, the PTF and OPC files reported an additional diagnosis per visit, resulting in a 19% higher estimate of diabetes prevalence. To address this problem, denominators used in the articles in this supplement include individuals 1) who were hospitalized and had diabetes as indicated by ICD-9-CM diagnosis code 250.xx and/or 2) who had three or more visits per year, with at least one of those being for diabetes. By using the three-visit standard, we hope to eliminate those infrequent users who may not have truly had diabetes or may have had the condition coded in error.

There are great advantages to using these national databases. First, once an investigator has obtained VA and human subjects' permission for data access, theoretically all VA inpatient and outpatient utilization can be tracked in files available on the Austin system. For investigators interested in the utilization of a specific group of patients, it is possible to upload a file of social security numbers to the mainframe and then merge this dataset with the files of interest to determine utilization. A second advantage is that current vital status information on a given set of veterans can be readily obtained. Third, when used with other databases, such as Medicare or hospital discharge abstracts from the states, it is possible to compare utilization in the VA with that in other systems. Examples include comparisons concerning the treatment of acute myocardial infarction in the VA and Medicare (6) and the use of coronary angioplasty in the VA and Washington State (11). Other advantages include the capacity to follow a cohort of patients over time and to capture more serious inpatient morbidity. In summary, these national files are an important resource for conducting epidemiologic, clinical, and health services research.

VA local databases

The VistA database contains extensive clinical information not available in the PTF or OPC files. The amount and type of information available in VistA exceeds that found in comparable nonfederal databases (12). Originating in the 1970s, VistA is a database management system that supports a variety of clinical, management, and system/database modules. Clinical modules cover dentistry, dietetics, laboratory, mental health, nursing,

pharmacy, radiology, oncology, social work, medicine, surgery, and quality management. In supporting the clinical functions, the management modules take care of patient scheduling and bed utilization, fiscal operations, file transfers to national databases, and eligibility queries. The system/database modules provide the interface between the computer operating system and the application modules, which have report-writing features. A key component of VistA is the Computerized Patient Record System (CPRS), which essentially reproduces the clinical chart in electronic format. The VA has been a leader in developing the computerized medical record that is now used in all VA medical centers.

Although the VistA database contains a large amount of clinical information, there are distinct limitations associated with its use. First, VistA does not have the capacity to do extensive data analysis. Extracting data in a format that can be analyzed using an external statistical software package can be difficult and complex. Second, although each VA medical center has a VistA, not all modules are available at each facility. Third, there are differences in VistAs across medical centers. Fourth, due to large amounts of information, medical centers routinely purge data at varying intervals. For example, in large centers laboratory or pharmacy data may be purged every 90 days; thus, if older records are needed, it may be necessary to go back to tapes. It is important to recognize that each VA has its own local system. If the desired information is available in one of the national databases, it is much easier to work with the Austin files than with >140 VistA databases.

Finally, to our knowledge, there is no published evaluation comparing VistA with medical records. An internal study done by the authors of volume 5 of *Department of Veterans Affairs Data Resource Guide* (12) compared information obtained from VistA with that obtained from medical records and other sources. In general, the utility of the VistA database was established, but there were discrepancies with respect to laboratory data. The authors caution that knowledge of the local system is needed to ensure that the information obtained is what was desired.

The CHIPS database is a data warehouse in use in VISN 20 that addresses many of the limitations of VistA. First,

since it is relational (Microsoft SQL Server), it is possible to join files from the various modules (e.g., laboratory and pharmacy). Second, given that extensive capabilities for file export exist in the database management software, exporting files for data analysis can be done with relative ease. The process of converting the *VistA* to the new SQL Server was both intensive and expensive and required several years to accomplish. Several VISNs have developed or are developing relational databases similar to CHIPS.

Other VA national databases

A complete description of all VA national databases is beyond the scope of this article. The VIREC website contains a catalogue of these databases and is the best source for current information on the many databases the VA maintains. A recent addition to the catalogue of VA databases is VetPop2000, the official estimate and projection of the number and characteristics of the veterans as of 30 September 2000, from the Office of the Actuary. This valuable resource resides on the VIREC website.

The DSS and the National Surgical Quality Improvement Program are two national databases that are of particular interest to epidemiologists and health services researchers. The DSS is a longitudinal database that combines selected elements of cost and clinical data from the PTF, OPC, and *VistA*. It utilizes proprietary software and provides data on patterns of care and treatment outcomes linked to resource consumption and health care costs. Each VA facility has a DSS site team and manager who are responsible for running reports requested by facility administrators, clinic managers, and others. Although the DSS is designed to support management decision making, it also has potential value for VA researchers. For researchers interested in cost issues, the VA Health Economics Research Center has created average cost data files for health care utilization for FY1998 through FY2001 (13).

The National Surgical Quality Improvement Program was created to provide reliable comparative data about the risks of mortality and morbidity in VA hospitals that perform major surgery and to gather accurate information about workload and length of stay. This initiative has its origins in the Center for Continuous Quality Improvement in Cardiac

Surgery, which has provided comparative data on morbidity, mortality, and workload for open-heart surgery in VA medical centers since the mid-1980s. Briefly, a surgical reviewer at each participating facility collects and transmits data to a coordinating center responsible for data management and analysis. Reports generated from the database include adjusted measures of mortality and morbidity for various surgical procedures performed in peer groups of VA hospitals.

Database access and patient confidentiality

In general, the PTF and OPC files are available only to VA researchers whose application to use them has been approved by both the local facility and the central office. These files contain scrambled social security numbers, preventing the identification of specific individuals. Researchers can also access files that contain patient identifiers, such as names and real social security numbers, although additional security clearance is required to use these files. Access to real social security numbers is needed in order to link the BIRLS death file to the PTF or OPC. Use of these files is a privilege that carries an obligation to treat them with due care and to implement and maintain procedures that ensure that data remain secure and confidential, according to standards of the Health Insurance Portability and Accountability Act of 2003. This means that data files should be stored on secure computers with access restricted to appropriate personnel. Data analysis files should have patient names, real social security numbers, and other personal identifiers removed, once all file linking has been accomplished. Permission to use VA local databases, such as *VistA* and CHIPS, is granted at the local or VISN level. Finally, the DSS and the National Surgical Quality Improvement Program database are generally not available to VA researchers. Investigators interested in using these databases should consult with the individual programs for more information.

New developments

At this time, significant changes in VA national databases are under way. The goals of these changes are to further enhance the National Patient Care Database by 1) eliminating fragmented and overlapping national data systems, 2) resolving inconsistencies among current national data

systems, 3) implementing data definitions consistent with international health care standards, 4) collecting full diagnostic and procedural information on outpatients, 5) shifting focus from service to patient, 6) improving timeliness of national data, and 7) taking advantage of modern technologies for information analysis. This process will take several years to complete, with the final product being a relational database advantageous to epidemiologists and clinical and health services researchers. Another significant effort on the horizon is the Health Data Repository, which is envisioned to be a longitudinal record of patient health information. This national registry is currently in the planning stages and will ultimately contain detailed health information similar to that in the *VistA* database. In addition to the services offered by VIREC, epidemiology research information centers provide technical assistance, as well as training and education, for VA researchers.

Conclusion

Both local and national VA databases are valuable resources for investigators who have a wide array of research interests. National data files such as the PTF and OPC are exhaustive in their coverage, yet lack important clinical detail. Local files, on the other hand, are more limited with respect to patient numbers and are difficult to use from a research perspective, yet have extensive clinical, laboratory, and pharmacologic data. The CHIPS database at the VISN level and eventually the Health Data Repository at the national level are relational databases that include laboratory and pharmacy data and are more friendly to researchers than the current cumbersome system. The possibilities for investigating important scientific questions with VA databases become greater as communications and database management technologies improve.

References

1. Department of Veterans Affairs, Veterans Information Research and Information Center [Web site]. Available at www.virec.research.med.va.gov
2. Ashton CM, Petersen NJ, Soucek J, Menke TJ, Yu HJ, Pietz K, Eigenbrodt ML, Barbour G, Kizer KW, Wray NP: Geographic variations in utilization rates in

- Veterans Affairs hospitals and clinics. *N Engl J Med* 340:32–39, 1999
3. Ashton CM, Petersen NJ, Wray NP, Yu HJ: The Veterans Affairs medical care system: hospital and clinic utilization statistics for 1994. *Med Care* 36:793–803, 1998
 4. Boyko EJ, Koepsell TD, Gaziano JM, Horner RD, Feussner JR: US Department of Veterans Affairs Medical Care System as a resource to epidemiologists. *Am J Epidemiol* 151:307–314, 2000
 5. Cowper DC, Kubal JD, Maynard C, Hynes DM: A primer and comparative review of major US mortality databases. *Ann Epidemiol* 12:462–468, 2002
 6. Fleming C, Fisher ES, Chang CH, Bubolz TA, Malenka DJ: Studying outcomes and hospital utilization for the elderly: the advantages of a merged database for Medicare and Veterans Affairs hospitals. *Med Care* 30:377–391, 1992
 7. Wright SM, Daley J, Fisher ES, Thibault GE: Where do elderly veterans obtain care for acute myocardial infarction: Department of Veterans Affairs or Medicare? *Health Serv Res* 31:739–754, 1997
 8. Page WF, Mahan CM, Kang HK: Vital status ascertainment through the files of the Department of Veterans Affairs and the Social Security Administration. *Ann Epidemiol* 6:102–109, 1996
 9. Dornitz JA, Maynard C, Boyko EJ: Assessment of vital status in Department of Veterans Affairs National Databases: comparison with state death certificates. *Ann Epidemiol* 11:281–286, 2001
 10. Kashner TM: Agreement between administrative files and written medical records: a case of the Department of Veterans Affairs. *Med Care* 36:1324–1336, 1998
 11. Ritchie JL, Maynard C, Chapko MK, Every NR, Martin DC: A comparison of percutaneous transluminal coronary angioplasty in the Department of Veterans Affairs and in the private sector in Washington State. *Am J Cardiol* 81:1094–1099, 1998
 12. Beattie MC, Swindle RW, Tomko LA, Moos RH, Recine BP, Polk DE, Adams TA: *Department of Veterans Affairs Data Resource Guide*. Vol. V. Indianapolis, IN, VA HSR&D Center for Health Care Evaluation, 1996
 13. Department of Veterans Affairs, Health Economics Research Center [Web site]. Available at www.herc.research.med.va.gov