Department of
Veterans Affairs

Development of an integrated stroke outcomes database within Veterans Health Administration

Dean M. Reker, PhD, RN;[1–2]* Kimberly Reid, MStat;[2–3] Pamela W. Duncan, PhD;[2–3] Clifford Marshall, MS;[4] Diane Cowper, PhD;[2–3] James Stansbury, PhD;[2–3] Kristen L. Warr-Wing, BA[2–3]

[1]*Kansas City Department of Veterans Affairs (VA) Medical Center, VA Rehabilitation Outcomes Research Center, and University of Kansas Medical Center, Kansas City, KS;* [2]*Health Services Research and Development (HSR&D) and Rehabilitation Research and Development, Rehabilitation Outcomes Research Center, North Florida/South Georgia Veterans Health System, Gainesville, FL;* [3]*Brooks Center for Rehabilitation Studies, University of Florida, Gainesville, FL;* [4]*Memphis VA Medical Center, Memphis, TN*

Abstract—A fundamental goal of the Rehabilitation Outcomes Research Center of Excellence is to improve care and outcomes for veterans with rehabilitation needs. To achieve this goal, the Center's primary objective is increasing research capacity. The Integrated Stroke Outcomes Database is a collection of Veterans Health Administration (VHA) clinical and administrative data containing patient information on a cohort of stroke patients found in the Functional Status Outcomes Database (FSOD), National Patient Care Database (NPCD), and other VHA sources. Clinical and administrative data were abstracted from several VHA data sources and linked to form an integrated outcomes database. A primary cohort of stroke patients treated during fiscal year (FY) 2001 was identified from the FSOD. Matching data from the NPCD, Decision Support System, Health Economics Resource Center, and the National Veterans Survey were obtained, merged, and reported in brief. This integrated database structure will provide valuable support to enhance the VHA capacity to perform stroke rehabilitation research.

Key words: costs, database, function, outcomes, rehabilitation, stroke, utilization.

INTRODUCTION

The Department of Veterans Affairs (VA) provides Veterans Health Administration (VHA) electronic data to VHA researchers at both the local hospital and national levels. These information systems have captured information on patients since the 1980s, providing an opportunity for researchers to conduct investigations on users of the VHA healthcare system over time. While the administrative data systems were not designed for research, researchers have used them to assess outcomes in numerous studies. The VHA operates and maintains the largest healthcare system in the United States. As of May 2003, VHA operated 160 hospitals, 134 nursing homes, 43 domiciliary units, and over 800 outpatient clinics under its purview [1]. The value of administrative data is that large-scale national-level studies can be conducted at relatively low cost. Extracting the same information from the patient's medical record would be prohibitively expensive, if not impossible. Some examples of studies that have used VHA administrative data for outcomes research can be found in Cowper et al. [2].

The Rehabilitation Outcomes Research Center (RORC) for Veterans with Central Nervous System Damage, founded in October 2001, is a new VHA Office of Research and Development Center of Excellence established jointly by the Health Services Research and Development and the Rehabilitation Research and Development branches. Its overall mission is to enhance access, quality, and efficiency of rehabilitation services through interdisciplinary research and dissemination activities, with a particular focus on stroke. Outcomes research capacity within the VHA is integral to this mission, and key steps in building such capacity are the assembling, linking, and integrating of the extensive VHA clinical and administrative data related to stroke.

The RORC's new Integrated Stroke Outcome Database (ISOD) is the embodiment of this vision and will be a key resource for outcomes researchers examining all phases of stroke care within the VHA. The ISOD joins patient information sources that are challenging to assemble yet potentially quite useful for improving the care of veteran stroke patients. Linking of these stroke outcome and administrative datasets provides researchers with a wealth of patient information they would not be able to access within one given dataset. Some of the valuable information the new linked datasets provide includes patient demographics, care setting, provider information, diagnostic codes, long-term care, outpatient treatment, and cost of treatment—all in one convenient central location. This database also allows researchers to follow patients' entire continuum of care throughout their stroke recovery. The remainder of this paper outlines the content and structure of the ISOD, presents basic descriptive statistics concerning VA stroke patients in the Functional Status Outcomes Database (FSOD), and familiarizes VHA health services and rehabilitation researchers with this important new data resource.

DESCRIPTION AND METHODS

The ISOD is a collection of VHA clinical and administrative data containing patient information on a cohort of stroke patients found in the FSOD [3], National Patient Care Database (NPCD) [4], and other VHA sources. Patient information in the ISOD consists of data components such as demographics, Functional Independence Measure (FIM) (reference) scores, procedure records, bed-section stay information, surgery records, outpatient visits, extended-care records, quality-of-life data, vital status, and costs of treatment. **Figure 1** depicts the structure of the ISOD.

The foundation of the ISOD is a cohort of patients who VHA clinicians identified as having a new stroke, evaluated using the FIM [5], and entered into the FSOD. The cohort, defined by codes of 1.1 through 1.9 (strokes) in the impairment group field of the FSOD, consists of 3,308 unique stroke patients with 3,588 patient admissions in FY2001. Healthcare services for these patients were provided in 182 unique treatment settings (some VHA facilities may have more than one treatment settings for stroke, i.e., acute rehabilitation unit, subacute rehabilitation unit, and nursing home). All stroke cases were selected if their rehabilitation discharge occurred during FY2001 (October 1, 2000, through September 30, 2001). The identification of stroke patients in the FSOD is considered the "gold standard" for a stroke diagnosis because a clinician establishes the stroke impairment code by chart review rather than by International Classification of Disease, 9th Revision (ICD-9), diagnosis codes, which can be notoriously unreliable [6].

Studies have reported the reliability and validity of the existing data sources used for this integrated database. Researchers have evaluated data from the NPCD across several variable domains [6–10]. FIM data have been
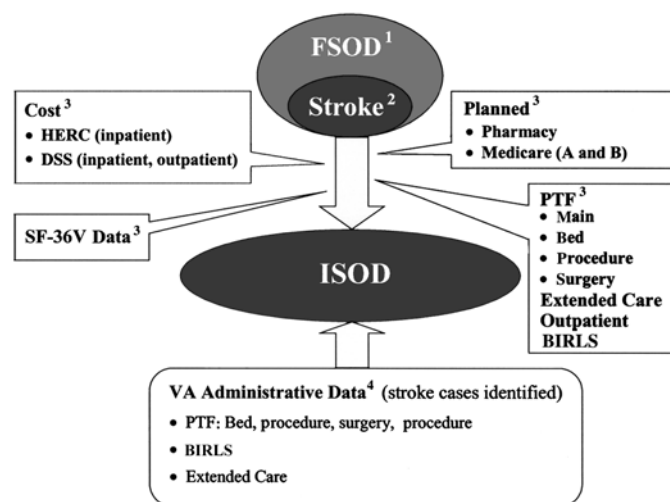
79

REKER et al. VHA integrated stroke outcomes database

Figure 1.

Integrated Stroke Outcomes Database (ISOD) for FY2001. 1. Parent database in which primary stroke cohort was identified. 2. Stroke cohort from Functional Status Outcomes Database (FSOD) was used as foundation for creating ISOD. 3. Data from these sources were obtained for patients in FSOD stroke cohort. 4. Additional stroke data were obtained from Department of Veterans Affairs (VA) administrative data with use of Reker's high-specificity, Reker's high-sensitivity, and Allocation Resource Center stroke algorithms. No corresponding cost of SF-36V data was obtained for these patients. BIRLS = Beneficiary Identification and Records Locator Subsystem, PTF = Patient Treatment File, HERC = Health Economics Resource Center, and DSS = Decision Support System.

extensively studied for instrument reliability across multiple settings [11–17]. Wagner et al. have recently examined the methodology of the Health Economics Resource Center (HERC) costing methods [18]. Finally, researchers at the RORC have assessed demographic data agreement between the FSOD and NPCD. This unpublished analysis compared date of birth, discharge destination, gender, marital status, and ethnicity/race. Agreement rates ranged from 96 to 99 percent with the exception of marital status (married 91%) and ethnicity/race (white 90%, African American 86%). Boehmer et al. and Stansbury et al. studied in detail additional ethnicity/race comparisons between the FSOD and NPCD [19–20].

In addition to the information on these stroke patients contained in the FSOD, we obtained further patient information from the NPCD and additional sources through unique patient identifiers. We obtained and merged all patient records associated with the primary FSOD stroke cohort into the ISOD. Sources of information included NPCDs: Patient Treatment File (PTF) main, bed section,

procedure, surgery, extended care, outpatient visits, Decision Support System (DSS) outpatient and discharge cost extracts, HERC outpatient and average cost discharge records [4], Beneficiary Identification and Records Locator Subsystem (BIRLS); and included SF (Short Form)-36V data from the Veterans Health Survey.

Because the FSOD has historically emphasized VHA stroke patients who have received formal rehabilitation services, the FSOD excludes a significant number of VHA stroke patients who have not received such services. Therefore, we identified additional stroke patients in the PTF main using several definitions based on ICD-9-CM (clinical modification) diagnostic codes. Although a VHA clinical directive mandated that all stroke, amputee, and traumatic brain injury patients be evaluated and listed on the FSOD beginning January 1, 2000, not all VHA stroke patients were identified and entered into the database. Because of this, additional tables establishing an independent patient sample were created that identify stroke patients with the use of ICD-9 codes and three ICD-9 diagnostic algorithms for stroke. The three algorithms used were (1) VHA Allocation Resource Center (ARC) definition for stroke [21], (2) a high-sensitivity*definition [8], and (3) high-specificity† definition [8].

Figure 2 illustrates membership in the differing ICD-9 stroke criteria during FY2001. As illustrated, patients and volume vary greatly, depending on the chosen ICD-9 definition of stroke. Once this large cohort of stroke patients was selected from the PTF main of the NPCD, all matching records from the procedure, bed section, surgery, extended care, and BIRLS files were obtained and added as additional tables (flat files).

Figure 2 represents 14,536 (10,681 + 1,644 + 2,211) patients meeting the ARC ICD-9 criteria, 9,577 (2,947 + 2,775 + 1,644 + 2,211) patients meeting the high sensitivity ICD-9 criteria, and 4,986 (2,775 + 2,211) patients meeting the high specificity ICD-9 algorithm.

Since the ISOD is a work in progress, all ISOD data just described are currently housed as distinct tables in SAS (Statistical Analysis Software) formats at the RORC at the Gainesville VA Medical Center (VAMC). The ultimate goal of the RORC is to construct this database

---

*High sensitivity captures most true stroke patients but also allows false-positive stroke patients into sample.

†High specificity limits the selection of false-positive stroke patients but does not capture a high proportion of true positive stroke patients.
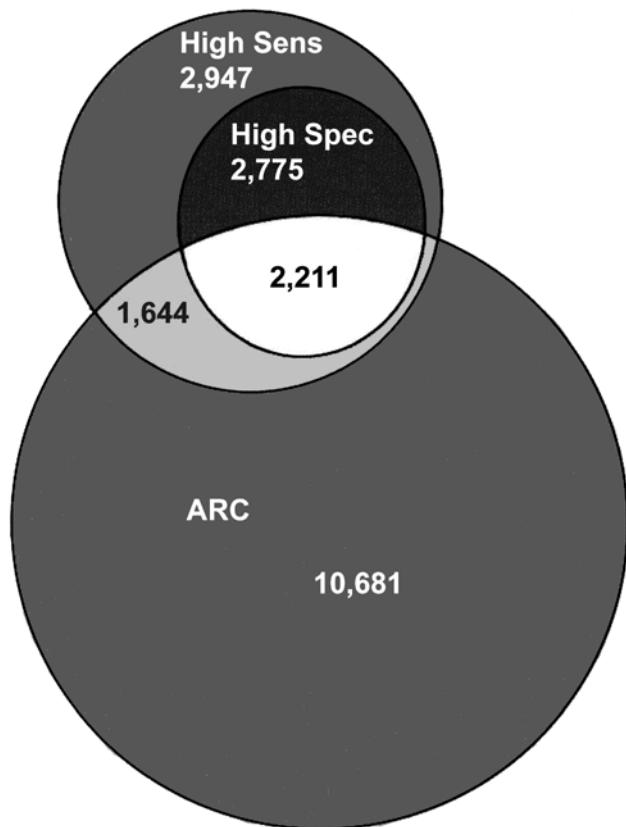
Figure 2.
Illustration of numbers of patients meeting three International Classification of Disease, 9th Revision, algorithms for stroke identification: high sensitivity (high sens), high specificity (high spec), and Allocation Resource Center (ARC).

annually and move it to the Austin Automation Center (Austin, Texas) or other VHA network server for use by clinicians and researchers throughout the VHA. With the ISOD linked to the FSOD in a seamless and integrated fashion, clinicians and researchers have real-time access to an array of previously unavailable information. Until this occurs, one can access the database by request through the RORC Web site [22].

We have generated descriptive statistics for the primary cohort of stroke patients obtained from the FSOD (**Table 1**). We have also generated additional descriptive statistics for each resulting analytic dataset following a matching and merging process with five additional unique data sources of the NPCD (**Tables 2–6**): PTF main, PTF extended care, PTF outpatient, DSS inpatient cost extracts, and DSS outpatient extracts. We also performed additional merges and data tables with HERC average cost discharge records, HERC outpatient, and SF-36V

Table 1.
Descriptive statistics from primary cohort of stroke patients abstracted from Functional Status Outcomes Database.

Variable	No. of Patients
Gender	
Male	3,478
Female	84
Total	3,562
Care Setting	
Acute Inpatient Rehabilitation	614
Subacute Inpatient Rehabilitation	194
Rehabilitation Continuum	2,780
Total	3,588
Marital	
Single	388
Married	1,679
Widowed	447
Separated	125
Divorced	803
Total	3,442
Ethnicity	
White	2,328
Black	858
Asian	22
Native American	10
Hispanic	288
Other	6
Admission Class	
Continuing Rehabilitation	142
Initial Rehabilitation	3,053
Readmission	85
Short-Stay Evaluation	134
Unplanned Discharge	69
Total	3,483
Functional Related Group	
ST-1	697
ST-2	273
ST-3	184
ST-4	389
ST-5	277
ST-6	290
ST-7	331
ST-8	187
ST-9	378
Total	3,006

ST = stroke

81

REKER et al. VHA integrated stroke outcomes database

Table 2.
Primary cohort match results with eight independent data tables.

| Variable | Primary Cohort | PTF | | | DSS | | HERC | | SF-36V |
		Main	Extended	Outpatient (SF File)	Inpatient	Outpatient	Discharge Record	Outpatient Record	
Gender									
Male (%)	98	98	98	98	98	98	98	98	98
Age (N)	(3,588)	(2,968)	(404)	(3,255)	(3,179)	(3,283)	(3,170)	(3,255)	(773)
Mean ± SD	68 ± 11	68 ± 11	69 ± 11	68 ± 11	68 ± 11	68 ± 11	68 ± 11	68 ± 11	72 ± 10
Median	70	70	70	70	70	70	70	70	74
[Min, Max]	[26, 97]	[26, 97]	[40, 97]	[26, 97]	[26, 97]	[26, 97]	[26, 97]	[26, 97]	[40, 91]
Length of Stay (d) (N)	(3,380)	(2,802)	(393)	(3,052)	(3,002)	(3,080)	(2,997)	(3,052)	(725)
Mean ± SD	25 ± 28	21 ± 24	34 ± 30	24 ± 27	22 ± 25	24 ± 27	22 ± 25	24 ± 27	23 ± 25
Median	17	15	27	16	16	16	16	16	16
[Min, Max]	[1, 295]	[1, 295]	[1, 290]	[1, 295]	[1, 295]	[1, 295]	[1, 295]	[1, 295]	[1, 184]
FIM Score									
Admission, Total (N)	(3,460)	(2,874)	(389)	(3,136)	(3,075)	(3,162)	(3,068)	(3,136)	(743)
Mean ± SD	72 ± 31	72 ± 31	67 ± 27	72 ± 32	71 ± 31	72 ± 32	71 ± 31	72 ± 32	72 ± 31
Median	74	74	69	74	73	74	73	74	74
[Min, Max]	[18, 126]	[18, 126]	[18, 124]	[18, 126]	[18, 126]	[18, 126]	[18, 126]	[18, 126]	[18, 126]
Admission, Motor (N)	(3,463)	(2,876)	(389)	(3,138)	(3,077)	(3,165)	(3,070)	(3,138)	(743)
Mean ± SD	48 ± 24	48 ± 24	44 ± 21	48 ± 24	48 ± 24	48 ± 24	48 ± 24	48 ± 24	48 ± 24
Median	48	48	42	48	47	47	47	48	48
[Min, Max]	[13, 91]	[13, 91]	[13, 91]	[13, 91]	[13, 91]	[13, 91]	[13, 91]	[13, 91]	[13, 91]
Discharge, Total (N)	(3,324)	(2,768)	(375)	(3,006)	(2,963)	(3,031)	(2,956)	(3,006)	(717)
Mean ± SD	90 ± 32	89 ± 32	90 ± 29	90 ± 32	89 ± 32	90 ± 32	89 ± 32	90 ± 32	89 ± 32
Median	100	99	100	100	99	100	99	100	100
[Min, Max]	[18, 126]	[18, 126]	[18, 126]	[18, 126]	[18, 126]	[18, 126]	[18, 126]	[18, 126]	[18, 126]
Discharge, Motor (N)	(3,325)	(2,769)	(375)	(3,007)	(2,964)	(3,032)	(2,957)	(3,007)	(717)
Mean ± SD	63 ± 24	63 ± 25	64 ± 23	63 ± 25	63 ± 25	63 ± 25	63 ± 25	63 ± 25	63 ± 25
Median	72	72	72	72	71	72	71	72	72
[Min, Max]	[13, 91]	[13, 91]	[13, 91]	[13, 91]	[13, 91]	[13, 91]	[13, 91]	[13, 91]	[13, 91]
Days Stroke Onset to Admission (N)	(3,588)	(2,968)	(404)	(3,255)	(3,179)	(3,283)	(3,170)	(3,255)	(773)
Mean ± SD	34 ± 269	29 ± 251	99 ± 504	35 ± 279	36 ± 283	35 ± 278	36 ± 283	35 ± 279	15 ± 69
Median	5	5	11	5	5	5	5	5	4
[Min, Max]	[0, 8,478]	[0, 8,478]	[0, 5,542]	[0, 8,478]	[0, 8,478]	[0, 8,478]	[0, 8,478]	[0, 8,478]	[0, 1,433]
Care Setting									
Acute Input Rehabilitation	614	584	36	478	598	598	596	479	118
Subacute Input Rehabilitation	194	155	123	148	193	193	193	148	37
Rehabilitation Continuum	2,780	2,229	245	2,629	2,388	2,388	2,381	2,628	606

DSS = Decision Support System   PTF = Patient Care File   SF-36V = veterans' health survey form
FIM = Functional Independence Measure   SD = standard deviation
HERC = Health Economics Resource Center   SF (short form) file = outpatient utilization data

Table 3.
Primary cohort merged with NPCD: PTF main and PTF extended care matches.

Variable	PTF Main (*N*)	PTF Extended Care (*N*)
Gender		
Male	2,881	392
Female	68	8
Total	2,949	400
Care Setting		
Acute Inpatient Rehabilitation	584	36
Subacute Inpatient Rehabilitation	155	123
Rehabilitation Continuum	2,229	245
Total	2,968	404
	% of Patients	
Mortality During Fiscal Year		
Survived year	92	89
Expired	8	11
Total	100	100
Source of Admission		
VA Hospital	—	54
Direct	60	—
Outpatient	30	—
Non-VA Hospital	4	—
Other	6	46
Total	100	100
Discharge Bed Section		
Rehabilitation	31	5
General Medicine	28	—
Neurology	21	—
Intermediate Medicine	15	—
Nursing Home	—	86
Geriatric Evaluation & Management Nursing Home	—	7
Domiciliary	—	1
Other	5	1
Total	100	100
Discharge Location		
Community	71	77
VA Nursing Home	15	9
Community Nursing Home	5	5
Expired	4	6
Other	5	3
Total	100	100

VA = Department of Veterans Affairs
NPCD = National Patient Care Database
PTF = Patient Treatment File

Table 4.
Primary cohort merged with NPCD: outpatient care files (SF file).

Variable	No. of Patients
Gender	
Male	3,159
Female	74
Total	3,232
Care Setting	
Acute Inpatient Rehabilitation	478
Subacute Inpatient Rehabilitation	148
Rehabilitation Continuum	2,629
Total	3,255
Insurance	
None	50
Medicare	37
Major Medical	6
PPO	3
Medicare Supplement	2
Other	2
Total	100
Prestroke Outpatient Visits[*]	
Laboratory	4,813
Primary Care—Medicine	3,271
Admission Screening	1,278
Ophthalmology	783
X-ray	637
Nursing	628
Other	14,398
Total	25,808
Poststroke Outpatient Visits[*]	
Laboratory	6,280
Primary Care—Medicine	3,989
Physical Therapy	2,608
Occupational Therapy	2,353
Admission Screening	1,951
Speech Pathology	1,987
Other	22,400
Total	41,568

[*]First clinic stop
NPCD = National Patient Care Database
SF = Short form (outpatient utilization data)
PPO = Preferred Provider Organization

83

REKER et al. VHA integrated stroke outcomes database

Table 5.
Primary cohort merged with NPCD: DSS inpatient cost extracts.

Variable	No. of Patients
Gender	
Male	3,087
Female	71
Total	3,158
Care Setting	
Acute Inpatient Rehabilitation	598
Subacute Inpatient Rehabilitation	193
Rehabilitation Continuum	2,388
Total	3,179
	% of Patients
Discharge Bed Section	
Rehabilitation Medicine	30
General Medicine	23
Neurology	17
Intermediate Medicine	13
Nursing Home	10
Other	7
Total	100
Primary Care Physician	
Physician: Internal Medicine	40
Physician: No Specialty	17
Nurse Practitioner	12
Physician Assistant	4
Other	27
Total	100
	Cost and Time
Inpatient Variable	
Total	$71.7 million
Cost Per Patient	$22,552
Inpatient Days	75,147
Average Length of Stay (d)	24
Average Cost Per Day	$954

NPCD = National Patient Care Database
DSS = Decision Support System

Table 6.
Primary cohort merged with NPCD: DSS outpatient cost extracts.

Variable	No. of Patients
Gender	
Male	3,087
Female	71
Total	3,158
Care Setting	
Acute Inpatient Rehabilitation	598
Subacute Inpatient Rehabilitation	193
Rehabilitation Continuum	2,388
Total	3,179
	% of Patients
Provider	
MD: Internal Medicine	12
MD: Resident	9
MD: No Speciality	9
Registered Nurse	5
Nurse Practitioner	4
Other	61
Total	100
	Cost
Prestroke Outpatient Stops	
Clinic Stops (*N*)	69,705
Mean $ Per Stop	131.89
Mean Pharmacy $ Per Stop	78.78
Mean Radiology $ Per Stop	235.64
Mean Laboratory $ Per Stop	69.57
Mean Surgery $ Per Stop	849.58
Mean All Other $ Per Stop	167.34
Poststroke Outpatient Stops	
Clinic Stops (*N*)	107,265
Mean $ Per Stop	160.22
Mean Pharmacy $ Per Stop	84.39
Mean Radiology $ Per Stop	254.58
Mean Laboratory $ Per Stop	63.57
Mean Surgery $ Per Stop	747.80
Mean All Other $ Per Stop	202.76

NPCD = National Patient Care Database
DSS = Decision Support System
MD = Medicinae Doctor (Doctor of Medicine)

data from the Veterans Health Survey (**Table 7**) [23]. **Table 2** displays descriptive variables from the primary cohort of stroke patients and the resulting sample descriptions after we matched and merged the eight data tables identified earlier in the Description and Methods section. Similarly, **Tables 3** through **6** report only on patient data from the matching table that corresponds with primary cohort of patients identified in **Table 1**.

Cost estimates in the DSS and HERC data are produced with the use of different costing methods. In brief, DSS cost data are based on a traditional healthcare accounting system and are calculated from indirect and direct costs at line-item product levels such as individual drugs, physical therapy visits, and consultations for individual patients.

Table 7.
Primary cohort merged with NPCD, HERC average cost discharge and outpatient records, and merged with SF-36 data from National Veterans Survey.

Variable	NPCD (HERC Records)		SF-36 Data
	Discharge	Outpatient	
Gender (*N*)			
Male	3,079	3,158	750
Female	71	74	19
Total	3,150	3,232	769
Care Setting			
Acute Inpatient Rehabilitation	596	479	120
Subacute Inpatient Rehabilitation	193	148	38
Rehabilitation Continuum	2,381	2,628	615
Total	3,170	3,255	773
Intensive Care Unit (ICU) (d)			
Total Days	3,161	—	—
Patients with ICU Stay	531	—	—
Average ICU Length of Stay	6	—	—
Rehabilitation-Related Costs[*]			
Patients with Specified Rehabilitation Costs[*]	$931.00	—	—
Total Cost	$25.8M	—	—
Cost Per Patient	$27,737.00	—	—
Inpatient Rehabilitation (d)	27,091	—	—
Average Rehabilitation Length of Stay (d)	29	—	—
Average Rehabilitation Cost Per Day	$953.00	—	—
Total Costs[*]			
Patients (*N*)	3,170	—	—
Total Cost	$75.8M	—	—
Cost Per Patient	$23,897.00	—	—
Inpatient Days	74,363	—	—
Average Length of Stay (d)	23	—	—
Category of Care (%)			
Diagnostic	—	30	—
Medicine	—	28	—
Rehabilitation	—	15	—
Surgery	—	8	—
Ancillary	—	6	—
Psychiatry	—	5	—
Other	—	8	—
Total	—	100	—
Prestroke Outpatient Stops			
Clinic Stops (*N*)	—	44,596	—
Mean National Costs[†]	—	$133.02	—
Mean Provider Costs[†]	—	$67.09	—
Mean Facility Costs[†]	—	$82.29	—
National Cost Estimate[†]	—	$109.38	—
Local Cost Estimate[†]	—	$107.52	—

85

REKER et al. VHA integrated stroke outcomes database

Table 7. (Continued)
Primary cohort merged with NPCD, HERC average cost discharge and outpatient records, and merged with SF-36 data from National Veterans Survey.

Variable	NPCD (HERC Records)		SF-36 Data
	Discharge	Outpatient	
Poststroke Outpatient Stops			
Clinic Stops (N)	—	72,054	—
Mean National Costs[†]	—	$125.53	—
Mean Provider Costs[†]	—	$67.28	—
Mean Facility Costs[†]	—	$70.76	—
National Cost Estimate[†]	—	$108.03	—
Local Cost Estimate[†]	—	$107.21	—
Bodily Pain (N = 796)			
Mean ± SD	—	—	43 ± 27
Median [Min, Max]	—	—	41 [0, 100]
General Health (N = 754)			
Mean ± SD	—	—	40 ± 22
Median [Min, Max]	—	—	38 [0, 100]
Mental Health (N = 766)			
Mean ± SD	—	—	62 ± 24
Median [Min, Max]	—	—	64 [0, 100]
Physical Functioning (N = 764)			
Mean ± SD	—	—	39 ± 28
Median [Min, Max]	—	—	35 [0, 100]
Role Emotional (N = 742)			
Mean ± SD	—	—	41 ± 48
Median [Min, Max]	—	—	20 [–17, 114]
Role Physical (N = 750)			
Mean ± SD	—	—	21 ± 37
Median [Min, Max]	—	—	2 [–7, 110]
Social Function (N = 766)			
Mean ± SD	—	—	52 ± 31
Median [Min, Max]	—	—	50 [0, 100]
Vitality (N = 769)			
Mean ± SD	—	—	37 ± 24
Median [Min, Max]	—	—	40 [0, 100]
Mental Summary Scale (N = 711)			
Mean ± SD	—	—	43 ± 13
Median [Min, Max]	—	—	43 [9, 72]
Physical Summary Scale (N = 711)			
Mean ± SD	—	—	31 ± 10
Median [Min, Max]	—	—	29 [9, 58]

*National case-mix adjusted costs
[†]Per clinic stop

SF-36 = short form for survey data
SD = standard deviation

In contrast, HERC cost data are synthetic estimates based on distributing facility costs to individual patients and are calculated in two ways with the use of statistical methods. For acute bed sections (internal medicine, neurology, etc.), costs are estimated based on a statistical cost function estimated from Medicare data and adjusted to reflect overall VHA cost experience. For an estimation of costs per stay for nonacute bed sections (intermediate care, nursing home, rehabilitation, etc.), costs per day for individual facility bed sections are multiplied by individual patient's lengths of stay. Additional documentation for DSS and HERC cost calculations can be obtained from the VHA Information Resource Center (VIReC) and HERC Web sites [4].

Planned data extractions and mergers with Medicare part A and B data, National Prosthetics Patient Database [4], and VHA Pharmacy (Pharmacy Benefits Management [database in NPCD] [4]) data have not yet occurred at the time of this printing.

We performed the dataset matching and merging process using SAS software (version 8.2) with the following basic logic. FSOD cases represented the primary dataset. We added other variables by matching first on primary patient identifier and then on assessment date. The initial FIM assessment date from the FSOD was required to occur between the admission and discharge dates found on the secondary inpatient data sources for a match and merge to occur. We matched and merged all outpatient records by a primary patient identifier and then categorized on the sequence of the visit, whether it was prestroke or poststroke during the fiscal year. We determined the stroke onset date from the FSOD database.

RESULTS

Tables 1 and 2 display descriptive information on the primary stroke cohort of patients that was extracted from the FSOD during FY2001. In this cohort, 3,588 stroke admissions were listed. Patients in the cohort were 98 percent male, mean age 68, and the median inpatient length of stay was 17 days, with a mean admission FIM score of 72 and a mean discharge FIM score of 90. The median period in days from stroke onset to rehabilitation admission was 5 days.

Seventeen percent (*n* = 614) of the primary cohort received rehabilitation care in an acute inpatient rehabilitation unit. These units provide the most intensive rehabilitation available in the VHA. Five percent of the cohort (*n* = 194) received rehabilitation care in subacute rehabilitation units (which also provide intensive rehabilitation services) but were typically housed in long-term care settings. The remainder of the cohort (77%) received care in other VHA settings, such as acute care beds, nursing home beds, or intermediate care beds. Severity of stroke as measured by functional-related groups (FRGs) [24] was fairly evenly distributed across the nine stroke FRGs with the exception of FRG-ST-1 (stroke-1—the most severe group) having a disproportional 23 percent of all cases.

Eighty-eight percent of admissions in the primary cohort completed their initial rehabilitation treatment, and seven percent of the admissions were either readmissions or continuing rehabilitation admissions. Sixty-six percent of admissions were ethnically classified as white, followed by 24 percent black, 8 percent Hispanic, and 2 percent other.

Table 2 reports descriptive variable information from the primary cohort database (FSOD) and the resulting sample characteristics after the matching and merging process with the additional data source table. As a direct result of the imperfect data merge process, the merged sample size decreases because not all patients in the primary cohort have matching data in the added table. This decrease in sample size may not always be random and, as a result, may introduce systematic bias in the merged data table. Hence, we created Table 2 to compare all the merged tables with the parent primary stroke cohort. For example, the mean and median ages across all data table merges are consistent at 68 and 70 years until the last column merges with the SF-36V data table in which the average age increases to 72 and median age to 74. Therefore, this increase in median age may indicate a potential source of bias for this matched sample, particularly since the sample size decreased from 3,588 down to 761. Additional observations on the matching characteristics will be discussed for each data source merge (Tables 3–7).

Table 3 displays descriptive statistics on admissions of patients from the primary cohort who were successfully matched and merged with the NPCD PTF main and PTF extended care. This merge allows researchers to examine patterns and variations in inpatient stroke care and rehabilitation. Since a perfect one-to-one match of the primary cohort and secondary data sources was not possible (probably caused by random key punch errors of social security numbers and admission dates), the descriptive means and median table, as well as the gender and care settings, are constructed exactly to match Table 1.

87

REKER et al. VHA integrated stroke outcomes database

The remaining frequencies of selected variables in each table were selected to provide examples of unique information available by joining of the secondary data source.

In **Table 3**, 83 percent ($n = 2,968$) of the primary cohort was successfully matched and merged with the secondary NPCD: PTF main dataset. The subset of matched cases in **Table 3** compares very closely across the common variables. Matches by care setting were, however, much higher for patients receiving acute rehabilitation care (95%) compared with subacute care (80%) and the continuum of care (80%). Additional information provided by the NPCD: PTF main secondary data source includes mortality during the fiscal year, source of admission, the bed section where the patient resided at discharge, and the discharge location.

Table 3 also displays successful matches with the use of the NPCD PTF extended care secondary source. The PTF extended care data source is almost identical in structure to the PTF main data source, except the patients in the extended care file receive their care in long-term care settings (nursing home, some intermediate care beds) and the patients in the PTF main receive their care in more acute care bed settings; however, the PTF main does have some rehabilitation, intermediate care, and nursing home bed sections.

Table 4 displays the results of the matching and merging of the primary admission cohort to the NPCD outpatient care file (SF file). Ninety-one percent of the primary admission cohort had matching outpatient records that occurred either pre- or poststroke. The descriptive statistics in the common fields were quite similar as in prior matches. New variable fields found in the outpatient files revealed information on insurance coverage and the departmental distribution of services pre- and poststroke. Of veterans in the primary admission, 50 percent cohort did not have any insurance benefits outside of the VA. Medicare covered 37 percent.

Outpatient data in the NPCD used for the ISOD are structured at the "visit" level (SF file), which represents a day that a veteran attends one or more clinics (stops) at the facility. Therefore, the data presented in **Table 4** are presented at the visit level, and the first clinic departmental stop for the veteran's visit is displayed for the pre- and poststroke visits. As expected, laboratory and primary care medicine accounts for 31 percent of first stop visits prestroke and 25 percent of first stop visits poststroke. Also as one may expect, poststroke visits were up 62 percent from the prestroke baseline of 25,808 visits, and the

distribution of poststroke first stop visit heavily favored physical therapy (6%), occupational therapy (6%), and speech pathology (5%).

Table 5 displays the results of the matching and merging process of the primary admission cohort to the NPCD DSS inpatient cost extract files. Eighty-nine percent of the primary admission cohort had matching inpatient cost files. The descriptive statistics of the resultant merged file were again comparable to the original cohort among the common variables. New information in the DSS cost extracts revealed the distribution of the primary care provider and inpatient costs associated with the inpatient episode of care. Approximately $72 million was spent for inpatient care for 3,179 admissions in the primary cohort. Average cost per inpatient admission was $22,552 with an average length of stay of 24 days.

Table 6 displays the results of the match merge of the primary admission cohort with the DSS outpatient cost extracts. As in prior tables, the merged subset file represented 91 percent of the original primary cohort with quite similar descriptive statistics. New information from the DSS outpatient secondary data source revealed the provider type in the outpatient setting and in the pre- and poststroke clinic stop costs. Since the DSS outpatient cost extracts are structured at the clinic stop level rather than the "visit" level, all data unique to **Table 6** will be presented at the clinic stop level. As observed in the visit level NPCD outpatient data, 54 percent more poststroke clinic stops were found compared with the prestroke baseline of 69,705 clinic stops. The average cost per clinic stop also increased from a prestroke average of $132 to a poststroke average of $160. Average clinic stop costs declined slightly for laboratory and surgery poststroke but increased slightly for pharmacy, radiology, and the "all other" category.

HERC average cost discharge records (inpatient) were merged with the primary stroke cohort as well (**Table 7**). This data table revealed similar results. Of the HERC cases, 88 percent were successfully matched and merged with the similar characteristics of the primary admission cohort. New information provided in this data table includes intensive care unit (ICU) days, rehabilitation-related costs, and total costs. Total inpatient HERC costs in **Table 7** are comparable with the total inpatient DSS costs in **Table 5**. In general, HERC costs are slightly higher in total costs (6% higher), cost per patient (+6%), and average cost per day (+7%).

Similarly, the HERC outpatient cost files are also shown in **Table 7**. The merge rates are slightly higher for

the HERC outpatient cost files compared with the HERC inpatient files (91% versus 88%). Merged cases, again, were very similar to the primary cohort. Similarities also exist between the HERC outpatient costs and DSS outpatient costs displayed in **Table 6**. The HERC outpatient files had considerably fewer outpatient stops because pharmacy stops are not included. Still, the mean cost per stop (prestroke only) was very similar to the DSS mean cost per stop ($133 versus $131).

Table 7 also represents the results of the merging of the primary stroke cohort with the SF-36V data from the 1999 National Veterans Survey. Of the 3,308 unique patients in the primary ISOD stroke cohort (23%), 773 were surveyed and responded to the National Veterans Survey during 1999. Respondent matches to the primary cohort averaged 4 years older than the entire cohort but were similar in length of stay, FIM measurements, onset days, gender proportions, and care settings.

SF-36V data, as part of the Veterans Health Survey in 1999, was collected approximately 2 years prestroke in this cohort. Respondents to the larger survey exceeded 850,000. SF-36V survey data collected in 1999 were matched with data for the patients in this FY2001 VA stroke cohort. The merged data table characterizes the VHA patient sample with much lower scores compared with national norms across all dimensions of the SF-36. Comparison scores (medians) of this group with the entire VHA sample[*] [25] ($N \approx 850,000$) are bodily pain, 41 versus 41; general health, 38 versus 45; mental health, 64 versus 68; physical functioning, 35 versus 55; role emotional, 20 versus 58; role physical, 2 versus 22; social function, 50 versus 63; and vitality, 40 versus 46. Based on these median comparisons, the matching cases with our primary stroke cohort have greater deficits than their fellow veterans, particularly in the physical dimension. Negative domain scores for "role emotional" and "role physical" are due to instrument changes (from the SF-36 to SF-36V) and scoring methods unique to the SF-36V.

Table 8 provides descriptive information on three additional cohorts of stroke patients meeting differing ICD-9 definitions for stroke. Similarities across the cohorts are observed for most descriptive variables. Population size is perhaps the most striking difference among

the groups with the largest cohort approximately three times the size of the smallest group. Bed section at hospital discharge and inhospital mortality also appear to vary among the groups.

DISCUSSION

This paper is an introduction to the ISOD that the RORC constructed at the North Florida/South Georgia Veterans Health System, Gainesville Division. The ISOD is a fundamental tool for investigating stroke rehabilitation outcomes, providing a case series based on a clinically diagnosed "gold standard" that will prove to be an invaluable tool for researching stroke rehabilitation in the VHA. The ISOD joins, in a single database, sources that were previously challenging to assemble. Thus it constitutes a national-level database that merges clinical, psychosocial, cost, and ultimately long-term outcomes (functional status across the continuum of care, subsequent mortality) in a single package, providing a tool of inestimable value for outcomes research in stroke, both from clinical and policy (health services) perspectives.

As a fully integrated stroke database, the ISOD is a resource for researchers interested in clinical (primary data collection) studies in stroke rehabilitation. The database allows for the generation of significant questions and hypotheses, based on a secure foundation of accurately designated cases. The strengths of this ISOD are—

1. The cohort of stroke patients is derived from a clinical database that identifies the patient (independently from ICD-9 codes) to have a stroke diagnosis.

2. The database combines multiple sources of data to allow for the identification of multiple independent and dependent variables, including physical function assessments.

3. The database combines both inpatient and outpatient services to identify and track services received near the index hospitalization of stroke.

4. The database contains cost data from two VHA sources for comparison: HERC average cost and DSS.

5. The database identifies stroke patients receiving the most intensive inpatient rehabilitation services that the VHA offers.

6. The database is recreated annually, thus allowing for trend analyses over time.

---

[*]Personal communication, Dr. Lewis Kazis (Bedford VA Medical Center), April 14, 2004.

89

REKER et al. VHA integrated stroke outcomes database

Table 8.
Descriptive information from three additional stroke cohorts meeting different ICD-9 definitions of stroke.

Variable	Allocation Resource Center	High Sensitivity	High Specificity
Patients (*N*)	14,705	9,670	4,989
Male (%)	98	98	98
Mean Age (yr)	69.5	68.1	68.1
Mean Length of Stay (d)	14.2	16.4	14.4
Married (%)	5	49	49
White (%)	63	64	64
1-Year Mortality (%)	20	20	21
Discharge Bed Section (%)			
Rehabilitation	5	13	8
General Medicine	49	34	37
Neurology	9	25	33
Intermediate Medicine	13	14	11
Discharge Location (%)			
Community	71	71	71
VA Nursing Home	13	10	11
Community Nursing Home	5	5	5
Expired	5	8	9

ICD-9 = International Classification of Disease, 9th Revision   VA = Department of Veterans Affairs

The weaknesses of the ISOD that are commonly associated with administrative databases are—

1. Limited studies on the reliability and validity of the data.
2. Inability to obtain perfect matches of data tables and the resulting shrinking sample sizes because of unmatched patient observations. These patient mismatches may be due to random key-punch errors when entering patient identifiers; however, one cannot completely rule out systematic patient mismatches and a resulting biased (nonrepresentative) patient sample.
3. The database does not capture 100 percent of all stroke patients in the VHA, and because of this, the cases in the ISOD may not represent the larger VA stroke population (selection bias).
4. Missing data within variable fields caused by data entry errors of omission. The range of missing data proportions is 0 to 31 percent; however, most variables have less than 5 percent missing values.

The improvement of the data quality in the ISOD depends on the capture rate for stroke patients in the FSOD. Since the VHA mandated FIM evaluations of stroke patients in the FSOD, the VHA followed with a national performance measure that quarterly measures the successful capture rate of stroke patients into the FSOD. This performance measure has steadily risen since its advent. This process measure will continue to improve the quality and completeness of the ISOD over time.

Given the emphases on local and regional specificities (or structures and process) as sources of variations in clinical practice and outcomes, the national scope of the data will prove to be extremely important. This scope of data not only can significantly enhance stroke rehabilitation in the VHA but also can serve research, increasing epidemiologic and health services research.

Data presented in this manuscript represent a portion of the available data fields within the ISOD. A data dictionary is available for the primary admission cohort taken from the FSOD. Additional documentation for the secondary sources is available at the VHA VIReC Web site [4]. SAS code for performing data mergers as done for these analyses is also available to VHA investigators. VHA investigators interested in acquiring the database for research can apply for a data-use agreement by contacting the RORC through their Web site [22].

CONCLUSION

The ISOD is an ongoing, annual compilation of administrative and clinical data forming a single source of comprehensive information on VHA stroke patients. The database has been structured to allow flexibility for case-selection criteria and case-matching rules to accommodate individual investigator needs. The scope of data availability covers inpatient and outpatient healthcare use, patient function, healthcare costs, and patient function. Because the ISOD is currently providing information to VHA stroke investigators, the goals of increasing research capacity for stroke and ultimately improving patient outcomes are being achieved.

ACKNOWLEDGMENT

REFERENCES

1. Planning Systems Support Group [homepage on the Internet]. Gainesville (FL): VHA Office of the Assistant Deputy Under Secretary for Health for Policy and Planning (ADUSH). VA Site Tracking System; May 2003 [cited 2003 Sep]. Available from: http://vaww.pssg.med.va.gov/.

2. Cowper DC, Hynes DM, Kubal JD, Murphy PA. Using administrative databases for outcomes research: Select examples from VA Health Services Research and Development. J Med Sys. 1999;23(3):249–59.

3. Physical Medicine and Rehabilitation, Department of Veterans Affairs: Functional Status Outcomes Database [database on the Internet]. Memphis [TN]: PMRS Central Office; June 1997 [upated 2004 Dec 10; cited 2003 Sep]. Available from: http://www1.va.gov/health/rehab/FSOD.htm/.

4. Veterans Affairs Information Resource Center [homepage on the Internet]. Hines [IL]: Department of Veterans Affairs (VA) Health Services Research and Development (HSR&D) Service [revised 2004 Dec 17; cited 2003 Sep]. Health Economics Resource Center [homepage on the Internet]. Menlo Park [CA]: Department of Veterans Affairs (VA) Health Services Research and Development (HSR&D) Service [updated 2004 Oct 26; cited 2003 Sep]. Available from: http://www.virec.research.med.va.gov/ and http://www.herc.research.med.va.gov/.

5. Granger CV, Hamilton BB. Linacre JM. Heinemann AW. Wright BD. Performance profiles of the functional independence measure. Am J Phys Med Rehabil. 1993;72(2): 84–89.

6. Kashner TM. Agreement between administrative files and written medical records: a case of the Department of Veterans Affairs. Med Care. 1998;36(9):1324–36.

7. Reker D, Hamilton B, Duncan P, Yeh S, Rosen A. Stroke: who's counting what? J Rehabil Res Dev. 2001;38(2): 281–89.

8. Cowper D, Kubal J, Maynard C, Hynes D. A primer and comparative review of major US mortality databases. Ann Epidemiol. 2002;12(7):462–68.

9. Murphy PA, Cowper DC, Seppala G, Stroupe KT, Hynes DM. Veterans Health Administration inpatient and outpatient data: an overview. Eff Clin Pract. 2002;5(3 Suppl):E4.

10. Hynes DM, Cowper D, Kerr M, Kubal J, Murphy PA. Database and informatics support for QUERI: current systems and future needs. Quality Enhancement Research Initiative [review]. Med Care. 2000;38(6 Suppl 1):I114–28.

11. Hsueh IP, Lin JH, Jeng JS, Hsieh CL. Comparison of the psychometric characteristics of the functional independence measure, 5 item Barthel Index, and 10 item Barthel Index in patients with stroke. J Neurol Neurosurg Psychiatry. 2002; 73(2):188–90.

12. Nelson DL, Melville LL, Wilkerson JD, Magness RA, Grech JL. Rosenberg JA. Interrater reliability, concurrent validity, responsiveness, and predictive validity of the Melville-Nelson Self-Care Assessment. Am J Occup Ther. 2002;56(1):51–59.

13. Cohen ME, Marino RJ. The tools of disability outcomes research functional status measures [review]. Arch Phys Med Rehabil. 2000;81(12 Suppl 2):S21–29.

14. Gosman-Hedstrom G, Svensson E. Parallel reliability of the functional independence measure and the Barthel ADL index. Disabil Rehabil. 2000;22(16):702–15.

15. Ottenbacher KJ, Hsu Y, Granger CV, Fiedler RC. The reliability of the functional independence measure: a quantitative review. Arch Phys Med Rehabil. 1996;77(12):1226–32.

16. Stineman MG, Shea JA, Jette A, Tassoni CJ, Ottenbacher KJ, Fiedler R, Granger CV. The Functional Independence Measure: tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. Arch Phys Med Rehabil. 1996;77(11):1101–8.

17. Vogel WB, Rittman M, Bradshaw P, Nissen D, Anderson L, Bates B, Marshall C. Outcomes from stroke rehabilitation in Veterans Affairs rehabilitation units: Detecting and correcting for selection bias. J Rehabil Res Dev. 2002; 39(3):367–84.

18. Wagner TH, Chen S, Barnett PG. Using average cost methods to estimate encounter-level cost for medical-surgical stays in the VA. Med Care Res Rev. 2003;60(3 Suppl):15–36S.

19. Boehmer U, Kressin NR, Berlowitz DR, Christiansen CL, Kazis LE, Jones JA. Self-reported vs administrative race/ethnicity data and study results. Am J Public Health. 2002; 92(9):1471–73.

20. Stansbury JP, Reid KJ, Reker DM, Duncan PW, Marshall CR, Rittman M. Why ethnic designation matters for stroke

91

REKER et al. VHA integrated stroke outcomes database

rehabilitation: Comparing VA administrative data and clinical records. J Rehabil Res Dev. 2004;41(3A):269–78.

21. VERA 2002 Patient Classification Chapter, VHA Allocation Resource Center, 100 Grandview Road, Suite 114 Braintree, MA.

22. Rehabilitation Outcomes Research Center [homepage on the Internet]. Gainesville [FL]: Department of Veterans Affairs (VA) Health Services Research and Development (HSR&D) and RR&D Center of Excellence [updated 2004 Oct 6; cited 2003 Sep]. Available from: http://www.vard.org/rorc/index.html/.

23. Kazis L. Short Form Health Survey for Veterans, Veterans Health Survey. VHA Office of Quality and Performance, http://vaww.oqp.med.va.gov/default.htm.

24. Stineman M, Ross R, Hamilton B, Maislin G, Bates B, Granger C, Asch D. Inpatient rehabilitation after stroke: a comparison of lengths of stay and outcomes in the Veterans Affairs and non-Veterans Affairs health care system. Med Care. 2001;39(2):123–37.

25. Health status and outcomes of veterans: Physical and mental component summary scores veterans SF-36, 1999 large health survey of veteran enrollees, executive report. Washington (DC): Department of Veterans Affairs Research and Development; May 2000.