

VA Health Economics Resource Center Technical Report #4

Reconciliation of DSS Encounter- Level National Data Extracts with the VA National Patient Care Database

Fiscal Year 2001

Wei Yu, PhD

Paul G. Barnett, PhD



*Dedicated to improving the quality
of health economics research*

Acknowledgements

The authors received valuable insights and comments from Elisabeth McSherry and Judith Garland at the Decision Support System Bedford Technical Support Office (BTSO), Denise Hynes at the VA Information Resource Center (VIREC), and Mark Smith at the Health Economics Resource Center (HERC).

Table of Contents

Executive Summary	v
Introduction	1
Inpatient Discharges	2
Merging variables and duplicate records	2
Results	2
Patterns of unmatched discharges	3
Recommendations to researchers	4
Files	7
Method	7
Results	8
Conclusions	9
Recommendations to researchers	9
Outpatient encounters	10
Distributions of DSS records and costs	10
Cost outliers	11
Sample selection	15
Duplicated visits	16
Outpatient encounter reconciliation	16
Recommendations to researchers	19
Recommendations to researchers	21
Discussion	22
Reference	22

Tables

Table 1. DSS NDE and NPCD PTF Inpatient discharge files FY2001	2
Table 2. Reconciliation of DSS NDE discharge file with PTF Main files (FY2001).....	2
Table 3. Frequencies of unmatched DSS discharges in the top 10 sites	3
Table 4. Frequencies of unmatched DSS discharges in the top 10 bedsections.....	3
Table 5. Frequencies of unmatched PTF discharges in the top 10 sites for each type of discharge (PM, PMO, XM)	5
Table 6. Frequencies of unmatched PTF discharges in the top 10 bedsection for each type of discharge.....	6
Inpatient Bedsection Stays	7
Table 7. DSS Treating Specialty and PTF Bedsection Files FY2001.....	7
Table 8. Equivalent File Names in PTF and NDE	8
Table 9. Reconciliation of DSS NDE Treating Specialty file with the three PTF bedsection files (FY2001)	9
Table 10. DSS National Outpatient Extract FY2001	10
Table 11. DSS Outpatient Clinic Extract records classification FY2001	11
Table 12. Number of outpatient records and total costs for each data source FY2001(all records)	11
Table 13. Cost outliers in the DSS Outpatient Extract (FY2001).....	12
Table 14. Cost outliers by STA3N in the DSS Outpatient Clinic Extract (FY2001).....	13
(Total cost >=\$100,000/single clinic encounter)	13
Table 15. Cost outliers for the top 10 clinic stops in the DSS Outpatient Clinic Extract (FY2001)	14
Table 16. Cost outliers by STA3N in the DSS Outpatient Pharmacy Extract (FY2001)	14
Table 17. A random sample of outpatient encounters in the DSS Outpatient Extract and NPCD Outpatient Event file (FY2001).....	15
Table 18. Reconciliation between the NPCD Outpatient Event file and the DSS Outpatient Extract in FY2001 (Sample)	16
Table 19. Telephone encounters in the unmatched NPCD Outpatient Event records FY2001	17
Table 20. Unmatched outpatient encounters for patients stayed in long-term care facilities by bedsection FY2001.....	18
Table 21. Comparison of DSS national extract discharge and treating specialty files	20
Table 22. The 5 medical centers with the most missing stays in the Discharge Extract.....	20

Executive Summary

We evaluated VA Decision Support System (DSS) National Data Extracts (NDEs) for FY2001.

Hospital discharge data. DSS discharge data and the VA National Care Patient Database (NPCD) discharge files had nearly identical information on inpatient stays. More than 99% of the records in the DSS data could be matched to a record in the NPCD. Most of the stays that appeared only in DSS had incomplete data; 85% did not include a value for the discharge bedsection. There were a small number of duplicate records in the NPCD.

Hospital bedsection stays. VA characterizes hospital stays by segments based on bedsection, the type of care provided according to the treating specialty of the physician. We compared the DSS inpatient treating specialty extract to NPCD bedsection files. We made data comparable by excluding DSS records of stays that were not over by the end of the year, by consolidating DSS data so that each stay in a bedsection was represented by one record, and by excluding stays in community nursing homes, which appear in NPCD but not DSS. About 40% of the bedsection stays in DSS did not have a corresponding stay in NPCD. We found a comparable percent of NPCD stays that had no match in the DSS file. The discrepancies were largely attributed to different methods of recording dates. Of the stays that differed in the two files, nearly half had a matching stay in the other database with a one-day difference in transfer date.

Outpatient cost outliers. The DSS pharmacy extract reports the cost of all drugs dispensed to an outpatient in a single day. There were 119 records with daily cost of more than \$100,000; 44% were from a single site. There were 252 other DSS outpatient encounters with a cost of more than \$100,000. Forty-four percent (44%) involved prosthetics services and were not confined to a single site.

Outpatient encounters. We compared records of outpatients in the DSS file that were flagged as having a corresponding encounter in the NPCD. Almost of all (99.9%) of the records in DSS had a corresponding encounter in the NPCD event file. About 7% of the records in the NPCD file did not have corresponding record in the DSS extract with the NPCD flag. About one-third of the differences were attributable to differences in the design of the two databases. After considering these differences, and liberal criteria for defining a match in the two sources, 3.5% of the total number of outpatient encounters in the NPCD had no corresponding DSS record.

Internal consistency of DSS data on hospital stays. We compared hospital stays that are represented in the two DSS inpatient files, stays that began and ended entirely within FY2001. We found 9 stays that were only in the discharge file and 4,575 stays that were only in the treating specialty file. We found discrepancy in costs reported in the two files: there were 153 stays with a difference in more than \$100, and 69 stays cost differences of more than \$1,000. There was better consistency in between these files than we found in FY2000.

Conclusion. DSS National Data Extracts provide useful information about VA utilization and cost at the encounter level. Researchers should always evaluate if data are complete and if cost estimates are reasonable. When an encounter is missing from DSS, its cost can be imputed using the average cost reported by DSS for a similar service.

Introduction

The Decision Support System (DSS) has been adopted by U.S. Department of Veterans Affairs to determine the cost of care provided in its nationwide network of hospitals and clinics. DSS cost data are produced by each medical center. To facilitate use of these data by headquarters staff, planners, and VA researchers, comprehensive encounter-level extracts of these DSS production data have been created and placed at the VA computer center in Austin, Texas. The computer files of these DSS extracts are quite large, with records representing each of the millions of patient care encounters provided each year to the nation's veterans.

Extracts from the VA National Patient Care Database (NPCD), also stored at the Austin computer center, include additional information about these encounters. This information includes patient characteristics, as well as the diagnoses and procedures associated with each encounter. These data are not repeated in the DSS extracts.

In order to associate DSS cost estimates with clinical and demographic characteristics of care, users of the DSS extracts must link DSS files to NPCD files. This report explains how the two databases can be matched. This reconciliation also has the value of understanding if DSS and NPCD completely characterize the services provided in the VA health care system.

HERC reconciled the two databases for the federal fiscal year ending on September 30, 2000 (FY2000) and reported the results in its research guide on the use of the DSS extracts¹. In the FY2000 reconciliation, we identified some discrepancies between the two databases, especially for outpatient care.

This report describes reconciliation of data from the two sources for FY2001. The reconciliation consists of three major parts: inpatient discharges, inpatient bedsections, and outpatient files. The inpatient discharge and treating specialty files in the DSS National Data Extracts (NDEs) were reconciled with the Patient Treatment Files (PTFs) in the NPCD database. The DSS Outpatient extract was reconciled with the NPCD Outpatient Event file (also called the SE file). For FY2001, the reconciliation method and results for inpatient care were very similar to that of FY2000. However, the reconciliation between the two outpatient databases was enhanced with more detailed investigation.

In addition to the reconciliation between the DSS NDE and the NPCD files, we also reconciled the DSS NDE treating specialty with the DSS discharge files. Compared with the result in FY00, the FY01 showed improvement in internal consistency of the DSS NDE inpatient files.

Inpatient Discharges

The DSS Discharge Extract includes all discharges that occurred during FY01 in VA facilities. However, inpatient discharges are separated among three PTF main files: the PM file for discharges from hospital main bedsections, the XM file for discharges from VA nursing homes, domiciliary and other residential health care facilities, and the PMO file for discharges from VA hospital observational bedsections. The numbers of discharges recorded in each file are listed in Table 1.

Table 1. DSS NDE and NPCD PTF Inpatient discharge files FY2001

Data source	Number of Records
DSS NDE discharge file	689,820
PTF discharge files	705,290
PTF main acute discharge file (PM)	566,318
PTF main non-acute discharge file (XM)	79,285
PTF main observation discharge file (PMO)	59,687

Merging variables and duplicate records

Before merging records in the two databases, a few of adjustments were made. First, discharges from community nursing homes were recorded in the PTF XM file, but not in the NDE discharge file. In FY2001, 11,141 discharges were recorded for community nursing homes in the XM file. After deleting those community nursing home discharges, 68,144 discharges remained in the XM file. Second, all the files were checked for duplicate discharge records.

The two databases were merged using the following four common variables: (1) scrambled Social Security Number (SCRSSN), (2) medical center identification number (3-digit numeric STA3N), (3) admission date (ADMITDAY), and (4) discharge date (DISDAY).

Records with the same values for the four variables listed above were considered to be duplicates. There were 206 duplicates in the PM file, three duplicates in the XM file and 58 duplicates in the PMO file. All duplicates were deleted before the merge.

Results

The reconciliation results of discharge files are summarized in Table 2.

Table 2. Reconciliation of DSS NDE discharge file with PTF Main files (FY2001)

	DSS NDE (n=689,820)	PTF Main (n=693,882)		
		PM (n=566,112)	XM (n=68,141)	PMO (n=59,629)
Merge with all four variables	684,210 (99.2%)	563,570 (99.6%)	66,763 (98.0%)	57,402 (96.3%)
Unmatched records	2,085	2,542	1,378	2,227

Table 2 indicates that over 99% of discharges in the two databases can be linked by these four variables. There are several possible explanations for the 1% of discrepancies between the two databases. First, some sites entered discharge data after costs had been allocated. Second, some long-term care patients stayed many years. DSS could not estimate costs for patients admitted before DSS was implemented; these hospital stays are not included in the DSS extracts. Third, DSS and PTF do not use the same rules to characterize observation care. Lastly, data entry errors contributed to the discrepancies. To identify these problems, unmatched records were further examined.

Patterns of unmatched discharges

There are two clear patterns among the records of hospitals stays that appeared in DSS but not in the PTF. First, while almost every site has a few examples of these stays, two sites accounted for 60% of these unmatched discharges. Table 3 lists the top 10 sites with unmatched discharges. Second, 85% of the unmatched discharges have missing values in the discharge bedsection. The frequencies of the top 10 bedsections (including the missing bedsection) for the unmatched discharges were summarized in Table 4.

Table 3. Frequencies of unmatched DSS discharges in the top 10 sites

STA3N	Frequency	Percent
Oklahoma City	822	39.83
NCHC Martinez	404	19.57
East Orange	111	5.38
Atlanta	96	4.65
N.Y. Harbor	51	2.47
San Juan	34	1.65
Long Beach	31	1.5
Upstate N.Y.	29	1.41
Miami	23	1.11
Jackson	23	1.11
SUM	1624	78.68

Table 4. Frequencies of unmatched DSS discharges in the top 10 bedsections

Bedsection	Frequency	Percent
Missing	1760	85.27
Gen (Acute) Med	75	3.63
Hi Int Gen Psch-Inp	45	2.18
Nursing Home Care	42	2.03
Intermediate Med	32	1.55
Sub Abuse Res Rehab	15	0.73
Star I, II, & III PGMS	11	0.53
Surgery (Gen)	10	0.48
Medical ICU	8	0.39
Spinal Cord Inj	8	0.39
SUM	2,006	97.19

We examined discharges in the PTF main files that did not have a matching record in the DSS discharge extract. We tabulated frequencies by site (STA3N) and bedsection for each of the three discharge files (i.e., PM, XM, PMO). Tables 5 and 6 report these statistics.

Table 5 shows that unmatched discharges are concentrated in a few sites, especially those for hospital discharges (PM discharges). The top 10 sites account for 72% of the unmatched PM discharges, 41% of the unmatched PMO discharges, and 55% of the unmatched XM discharges. These unmatched discharges are also narrowed in the top 10 bedsections, accounting for 86% of all the unmatched PM discharges (see Table 6). Nursing home and domiciliary account for 90% of the unmatched XM discharges. (Note that 1,760 discharges in the DSS NDE have missing values in the discharge bedsections. It is possible that these discharges do not have complete data.)

Recommendations to researchers

Overall, the discharge files in the two databases reconciled fairly well. Researchers should be able to link clinical information with DSS costs for VA inpatient care at a reasonable level of accuracy. Although the discharge files reconciled very well, researchers should always compare records in the two databases. For any stays missed by the DSS, costs can be imputed using the average costs of similar hospital stays. Researchers should also check for integrated stations because the two databases may change station number for an integrated station at different time. Information on integrated stations can be obtained from the Deputy Under Secretary for Health for Operations and Management website (<http://vhacoweb1.cio.med.va.gov/adush>).

Table 5. Frequencies of unmatched PTF discharges in the top 10 sites for each type of discharge (PM, PMO, XM)

Type	Site (STA3N)	Frequency	Percent
PM	Tampa	579	22.84
PM	Asheville-Oteen	282	11.12
PM	Illiana HCS Danville IL	258	10.18
PM	Colmery-Oneil VAMC <200107	182	7.18
PM	San Juan	159	6.27
PM	Murfreesboro	100	3.94
PM	Upstate N.Y.	69	2.72
PM	Tuscaloosa	68	2.68
PM	Iowa City <1000	66	2.6
PM	East Orange	55	2.17
PM	SUM	1818	71.7
PMO	Houston	253	11.41
PMO	Upstate N.Y.	107	4.83
PMO	Louisville	85	3.83
PMO	Tampa	74	3.34
PMO	Beckley	69	3.11
PMO	W Palm Beach	69	3.11
PMO	N.Y. Harbor	68	3.07
PMO	Detroit	62	2.8
PMO	VAMC Wichita KS <102001	59	2.66
PMO	Dallas	57	2.57
PMO	SUM	903	40.73
XM	Colmery-Oneil VAMC <200107	306	22.3
XM	Murfreesboro	84	6.12
XM	New Orleans	82	5.98
XM	White City	69	5.03
XM	Asheville-Oteen	46	3.35
XM	Sth Colorado HCS<1001	41	2.99
XM	Poplar Bluff<401	37	2.7
XM	Columbia MO<0401	35	2.55
XM	East Orange	29	2.11
XM	Dallas	25	1.82
XM	SUM	754	54.95

Table 6. Frequencies of unmatched PTF discharges in the top 10 bedsection for each type of discharge

Type	Bedsection (DBEDSECT)	Frequency	Percent
PM	Gen (Acute) Med	850	33.53
PM	Hi Int Gen Psch-Inp	262	10.34
PM	Intermediate Med	224	8.84
PM	Psych-General Inter	172	6.79
PM	Star I,II,&III PGMS	170	6.71
PM	Cardiology	161	6.35
PM	Surgery (Gen)	129	5.09
PM	Medical ICU	112	4.42
PM	Pulm Non-TB	52	2.05
PM	Orthopedic	50	1.97
PM	SUM	2182	86.09
PMO	Medical Observation	1538	69.37
PMO	Surgical Obs	373	16.82
PMO	Psychiatric Obs	285	12.86
PMO	Neurology Obs	14	0.63
PMO	SCI Observation	7	0.32
PMO	SUM	2217	100
XM	Nursing Home Care	923	67.27
XM	Domiciliary	312	22.74
XM	Sub Abuse Res Rehab	48	3.5
XM	Psyc Res Rehab Trmt	32	2.33
XM	Dom Substance Abuse	17	1.24
XM	Gem NHCU	8	0.58
XM	HCMC CTW/TR	7	0.51
XM	Dom Care Hmls(DHCV)	4	0.29
XM	Intermediate Med	3	0.22
XM	Medical ICU	2	0.15
XM	SUM	1356	98.83

Inpatient Bedsection Stays

Files

Records in the NDE treating specialty file were matched with those in the PTF bedsection files. The NDE Treating Specialty extract is a single file, while PTF Bedsection records are stored in three files: the Acute Bedsection file (PB), the Non-acute (or “Extended”) Bedsection file (XB), and the Observation Bedsection file (PBO). File sizes are listed in Table 7.

Table 7. DSS Treating Specialty and PTF Bedsection Files FY2001

Source	Number of Records
DSS NDE Treating Specialty file	1,293,596
PTF Bedsection files, total	907,460
PTF Acute bedsection file (PB)	765,099
PTF Non-acute bedsection file (XB)	82,620
PTF Observation bedsection file (PBO)	59,741

Method

Prior to merging NDE and PTF files, several steps were taken to modify them so that they recorded hospital care in the same format. This section details these steps.

The NDE Treating Specialty file contains bedsection stays that were not discharged or transferred, known as “Census stays.” The PTF bedsection files do not include records of these Census stays. Census stay records in the NDE Treating Specialty extract were excluded from further analysis by eliminating all records with a value of “Y” for the census stay variable (CENSUS). Although there is a PTF census file, this study did not attempt to reconcile it to the census records in DSS treating specialty file.

Next, monthly records in the DSS treating specialty file were consolidated into one record for each unique bedsection stay. As discussed in Chapter 6 of the Research Guide¹, the NDE Treating Specialty extract contains cost information for each bedsection stay by monthly fiscal period. If a bedsection stay lasts more than a single fiscal period, there will be multiple records for the same stay. These multiple records have the same values for five variables (SCRSSN, STA3N, TRTIN, TRTOUT, and TRTSP). The treating specialty file can be consolidated using these variables.

Since the DSS NDE Treating Specialty extract did not contain data from community nursing homes, we excluded community nursing home stays from the PTF XB file by eliminating records with “STATYP=42” (the variable STATYP is kept in the main PTF extended care file; thus the main and bedsection files were merged to identify which records to exclude). Records with the same values in the five merging variables were considered as duplicates and removed before the merge. The following shows the results of the file merge:

NDE Treating Specialty extract (FY01):	1,293,596
Census Records:	-156,728
Consolidated multiple-record stays:	- 248,190
<i>NDE Treating Specialty records for the merger:</i>	<u>888,678</u>
PTF Main Bedsection file (PB01)	765,099
duplicates:	-1,078
PTF PB records for the merger:	<u>764,021</u>
PTF Extended Bedsection file (XB01):	82,620
Community Nursing Homes:	-11,145
Duplicates:	-28
PTF XB records for the merger:	<u>71,447</u>
PTF observational bedsection (PBO01)	59,741
duplicates:	-91
PTF PBO records for the merger:	<u>59,650</u>
<i>Total PTF bedsection records for the merger:</i>	<u>895,118</u>

The DSS NDE and PTF bedsection files were merged by the following five variables: (1) scrambled Social Security Number (SCRSSN), (2) medical center identification number (3-digit numeric STA3N), (3) bedsection admission date, (4) bedsection discharge or transfer date, and (5) bedsection number.

Please note that three variables have different names in the two databases. The bedsection admission date is called BSINDAY in the PTF and should have the same value as TRTIN in the DSS treating specialty file. The bedsection discharge or transfer date is called BSOUTDAY in the PTF and should have the same value as TRTOUT in the DSS treating specialty file. Bedsection number is called BEDSECN in the PTF and should have the same value as TRTSP in the DSS file. Table 8 lists the equivalent variable in the two sources.

Table 8. Equivalent File Names in PTF and NDE

PTF File Names	NDE File Names
BSINDAY	TRTIN
BSOUTDAY	TRTOUT
BEDSECN	TRTSP

Results

Nearly 40% of the records in each file were not matched. Results are summarized in Table 9. The main problem of matching bedsection stays between the two databases is that the DSS system defines a day of stay differently than the NPCD. The DSS system does not record a day if the patient stay in the bedsection is less than 24 hours, however the NPCD does. The detailed patterns of these bedsection discrepancies were examined in FY00 data and reported in Research Guide to the DSS National Data Extracts ¹.

Table 9. Reconciliation of DSS NDE Treating Specialty file with the three PTF bedsection files (FY2001)

	DSS NDE (n=888,678)	PTF Bedsection Files (n=895,118)		
		PB (n=764,021)	XB (n=71,447)	PBO (n=59,650)
Merge with all five variables	651,521	539,888	53,816	57,817
Non-matched records	237,157 (36.4%)	224,133 (41.5%)	17,631 (32.8%)	1,833 (3.1%)

Conclusions

The patterns of reconciliation between the DSS Treating Specialty Extract and the PTF Bedsection files in FY2001 were very similar to those in FY2000. Nearly 40% of bedsection stays were not matched by the five linking variables. This was mainly caused by the different criterion of crediting a day of inpatient stay between the two systems. Most of the inconsistencies occurred on the date recorded for bedsection transfer. In general, the DSS bedsection transfer date is one day less than the PTF bedsection transfer date.

Recommendations to researchers

If a study needs to identify bedsection stays, researchers should first extract all bedsection stays from the NPCD and DSS databases for the study subjects using scrambled social security numbers, regardless of other information. Then, researchers can adopt the method used in this section to conduct the first match. For the unmatched bedsection stays, researchers are recommended to conduct further matches in four steps. First, drop the 3-digit station number and match the records by scrambled SSN, bedsection in and out dates, and bedsection name. This can identify any inconsistencies caused by medical center integrations in the middle of a fiscal year. Researchers need to obtain information about integrated medical centers during the year and check that the resulting match makes sense. Information on integrated stations can be obtained from the Deputy Under Secretary for Health for Operations and Management website (<http://vhacoweb1.cio.med.va.gov/adush>). Second, drop the bedsection-out date and match the rest of the records by station number, scrambled SSN, bedsection-in date, and bedsection name. About half of the unmatched records are expected to be matched in this step. Third, drop the bedsection-in date and match the records by the other four matching variables. Finally, researchers should sort the still unmatched records by station number, scrambled SSN, and bedsection name, print out these unmatched records, and examine manually for any possible match. Since the numbers of stays recorded in both databases are very close, researchers should be able to match most of the bedsection stays.

Outpatient encounters

The method of reconciling outpatient data in the DSS extracts and NPCD was similar to the methods employed for FY2000. The method was enhanced by increasing the sample size to 10% of outpatients, ten times the 1% sample used in FY2000. Additional analyses not conducted in FY2001 were undertaken. The DSS outpatient extracts exclude telephone consultations and care the outpatient clinics provide to domiciliary (it is considered inpatient care by DSS). The FY2001 analysis examined how important these problems were in explaining discrepancies between NPCD and DSS outpatient data.

The entire DSS dataset was evaluated, including care reported in DSS that is not included in the NPCD. Cost outliers were identified from the entire DSS outpatient extract, including both clinic and pharmacy extracts.

Distributions of DSS records and costs

The DSS Outpatient extract consists of two files: the outpatient pharmacy file and the outpatient clinic file. The outpatient pharmacy extract contains encounters at the VA pharmacy clinic and the outpatient clinic extract includes all other outpatient encounters. The number of records and total costs are listed in Table 10.

Table 10. DSS National Outpatient Extract FY2001

File	Total number of records	Total costs
Outpatient Clinic	63,095,782	\$5,211,636,542
Outpatient Pharmacy	47,961,983	\$3,061,169,153

Because the NPCD database does not include outpatient pharmacy records except consultations provided at pharmacy clinics, we excluded outpatient pharmacy records for the reconciliation. The DSS system catches all outpatient activities recorded in the VistA system and many of these activities were not recorded in the NPCD database. To identify those activities, the DSS outpatient extract has eight categories classified by seven flag variables, each representing a primary data source, and an eighth category, which has none of the flags. There is also a combined flag variable called ENCFLAG that summarizes the eight categories. Table 11 describes the eight categories and Table 12 summarizes the number of records and costs for each category in the FY2001 clinic outpatient extract.

Table 11. DSS Outpatient Clinic Extract records classification FY2001

Flag Variable	Description
PRE	Outpatient pharmacy clinic encounter records
NOS	Outpatient clinic no-show records
PRO	Records extracted from the VistA prosthetics package
DDC	Records extracted from the Denver Distribution Center
CLI	Encounter records extracted from VISTA and not overwritten by the NPCD records
UTIL	When no encounter records can be found with which to link outpatient utilization records from the SUR, LAB, RAD, ECS and ECQ feeder systems, a separate Util-Built encounter is created for each SSN, Date, and Stop code combination.
NPCD	Records from the Austin NPCD outpatient data collection system. NPCD records overwrite the CLI records.
None	<p>“All Other” type of care, including</p> <ul style="list-style-type: none"> • Mental health testing (clinic stop 538) • Addiction Severity Index assessment • Community nursing home, state nursing home, state domiciliary, state hospital (clinic stops 650, 651, 652, 653) • Contract homeless, alcohol/substance abuse & HCMI (clinic stop 654 in VISN 22 only) • Purchased home care (Clinic stop 681) • No stop code - utilization records not otherwise posted to established encounters (SSN = 100101000)

Table 12. Number of outpatient records and total costs for each data source FY2001 (all records)

Category	% of Total Records (N=63,095,782)	% of Total Cost (Total Cost=\$5,211,636,542)
NPCD	80.47	70.12
CLI	2.26	1.80
PROS	3.39	4.33
DDC	0.73	1.38
NOSHOW	7.17	4.27
UTIL	3.94	7.30
None	1.48	9.90
MULTIPLE	0.29	0.44

It is clear that DSS captures much more cost activities for outpatient services than the NPCD database does. Among them, records listed in the CLI and UTIL account for 9.2% of the total costs, records in NOFLAG account for 9.9% of total costs.

Cost outliers

Records with total cost of \$100,000 or higher for a single clinical encounter were identified from both the DSS outpatient clinic and pharmacy extracts. We listed these outliers in Table 13.

Table 13. Cost outliers in the DSS Outpatient Extract (FY2001)

(Total cost >= \$100,000/single clinic encounter)

	Pharmacy	Other Clinics
Number of records	119	252
Total cost	\$27,290,686	\$31,308,888
Maximum cost / single stop visit	\$860,167	\$866,797

These cost outliers are further examined by medical center (STA3N) and clinic stop (for clinic outliers only) in Table 14 through Table 16.

Table 14. Cost outliers by STA3N in the DSS Outpatient Clinic Extract (FY2001)
(Total cost >=\$100,000/single clinic encounter)

STA3N	STA3N Label	Frequency	Percent	Cumulative Frequency
618	Minneapolis	15	9.87	15
541	Cleveland-Wade Park	14	9.21	29
520	Gulf Coast HCS	11	7.24	40
521	Birmingham	10	6.58	50
580	Houston	10	6.58	60
630	N.Y. Harbor HCS	9	5.92	69
619	Montgomery	7	4.61	76
512	Baltimore	6	3.95	82
671	San Antonio	6	3.95	88
459	Honolulu	5	3.29	93
629	New Orleans	5	3.29	98
528	Upstate N.Y.	4	2.63	102
554	Denver	4	2.63	106
578	Hines	4	2.63	110
660	Salt Lake City Hthcare	4	2.63	114
537	VA Chicago HCS	3	1.97	117
598	Little Rock	3	1.97	120
523	Boston	2	1.32	122
526	Bronx	2	1.32	124
595	Lebanon	2	1.32	126
612	NCHC Martinez	2	1.32	128
657	VA Heartland-E VH MO	2	1.32	130
672	San Juan PR	2	1.32	132
677	Colmery-ONeil VAMC<200107	2	1.32	134
689	West Haven	2	1.32	136
504	Amarillo HCS	1	0.66	137
508	Atlanta	1	0.66	138
529	Butler	1	0.66	139
534	Charleston	1	0.66	140
539	Cincinnati	1	0.66	141
546	Miami	1	0.66	142
552	Dayton	1	0.66	143
565	Fayetteville NC	1	0.66	144
568	Fort Meade	1	0.66	145
596	Lexington-Leestown	1	0.66	146
600	VA Long Beach HCS CA	1	0.66	147
605	Loma Linda	1	0.66	148
608	Manchester	1	0.66	149
637	Asheville-Oteen	1	0.66	150
652	Richmond	1	0.66	151
663	Puget Sound HCS	1	0.66	152

Table 15. Cost outliers for the top 10 clinic stops in the DSS Outpatient Clinic Extract (FY2001)

(Total cost >=\$100,000/single clinic encounter)

CL	CL Label	Frequency	Percent	Cumulative Frequency
423	Prosthetics	67	44.08	67
146	PET	14	9.21	81
125	Social work services	7	4.61	88
429	Outpatient care in the O.R. room	7	4.61	95
553	Day treatment group	7	4.61	102
UNK		5	3.29	107
323	Primary Care/Med	5	3.29	112
650	Unknown	4	2.63	116
105	X-Ray	3	1.97	119
108	Laboratory	3	1.97	122

Table 16. Cost outliers by STA3N in the DSS Outpatient Pharmacy Extract (FY2001)

(Total cost >=\$100,000/single clinic encounter)

STA3N	STA3N Label	Frequency	Percent	Cumulative Frequency
630	N.Y. Harbor	52	43.7	52
405	White River Jct	11	9.24	63
452	VAMC Wichita KS<102001	10	8.4	73
528	Upstate N.Y.	8	6.72	81
553	Detroit VAMC	7	5.88	88
642	Philadelphia	4	3.36	92
629	New Orleans	3	2.52	95
657	VA Heartland-E VH MO	3	2.52	98
662	San Francisco	3	2.52	101
436	Fort Harrison	2	1.68	103
501	New Mexico	2	1.68	105
593	Las Vegas	2	1.68	107
609	Marion IL<0701	2	1.68	109
620	Montrose VA Hudson HCS NY	2	1.68	111
652	Richmond	2	1.68	113
678	Southern Arizona	2	1.68	115
538	Chillicothe	1	0.84	116
540	Clarksburg	1	0.84	117
589	VAMC Heartland-W Kansas MO	1	0.84	118
598	Little Rock	1	0.84	119

Sample selection

Because the number of outpatient records was very large, we selected a 10% random sample of VA patients for the reconciliation. We extracted all the outpatient records from the two databases for people whose last two digits of the scrambled Social Security Numbers were between 60 and 63 (including 60 and 63).

The NPCD event file contains all the clinical encounters incurred by a patient; the care of some patients is represented by multiple records for a single clinic stops on a single day. The DSS Outpatient Extract is designed to use one record to represent all care provided in a single clinic stop on a single day. In order to compare the two sources, the NPCD Event file were consolidated so that there was one record for each single clinic stop visited by a patient on a given day, the same rule applied in creation of DSS data. The DSS data were limited to records with the NPCD flag.

Table 17 compares data on the sample cohort from the two sources after these adjustments.

Table 17. A random sample of outpatient encounters in the DSS Outpatient Extract and NPCD Outpatient Event file (FY2001)

Data Source	Number of People	Number of Records
DSS Outpatient Clinic Extract	306,057	4,729,627
NPCD Outpatient Event File	310,102	5,653,059

Table 17 shows that not only the number of records in the NPCD outpatient event file is larger than that in the DSS outpatient clinic extract; the number of people is also larger in the NPCD Event file. There were 3,203 people who appeared only in the NPCD event file and 163 people who appeared remained only in the DSS Outpatient Extract.

The next analysis determined whether any of the 3,203 people who appeared only in the NPCD file might have records in the DSS pharmacy extract, or records in the DSS clinic extract that did not have the NPCD flag. There were still 1,758 people in the NPCD outpatient files who had no records in either of the DSS outpatient files. The DSS Outpatient Extract does not include any care received by these 1,758 people in FY2001. For the entire VA system in FY2001, the estimated number of people who were missed by the DSS NDE is about 17,000 (0.6%).

In order to understand the characteristics of people who were completely missed by the DSS NDE, we examined all records in the original files from a single VISN (Veterans Integrated Service Network). We found that the VISN had 185,437 people recorded in the NPCD outpatient event file in FY2001, of which 1,791 (1%) were not recorded in any of the DSS NDE outpatient files. The percentage of missed people by the DSS NDE in the single VISN was consistent with that found in the sample. Among the 1,791 persons, 746 (42%) had only telephone services that were not supposed to be included by the DSS NDE because those services are not assigned costs separately by the DSS. This left 1,045 people who were truly missed by the DSS NDE. However, 1,045 missed people had only 1,233 encounters in FY2001, indicating that most of them had only one outpatient encounter during the year. This suggests

that the DSS NDE missed some outpatient records. When those records were the only records for some patients, they were completely missed by the NDE.

Duplicated visits

The DSS NDE outpatient file contains one record for each clinic encounter in a single visit. If a patient had two or more visits to the same clinic during a given day, these encounters would be consolidated into a single record. We checked in the sample to see if there were any duplicated records for visits to the same clinic on the same day. We found that in the NPCD category (flag NPCD="Y"), there were no duplicates. However, there were 102 duplicates between the NPCD category and other categories. Most duplicates had the CLI and UTIL flags. We also found 165 duplicates under the no-flag category. Except for two records, all of the duplicated encounters were from the Psychology testing clinic (CL=538). A total of 267 total duplicates is trivial considering that there were 4.7 million records in the sample. These duplicates should cause few problems for researchers when linking the two databases, although they must be considered when merging the two data sources.

Outpatient encounter reconciliation

After consolidated duplicated clinic encounters incurred by a patient during a single visit, the two samples were merged by the following four variables: (1) scrambled Social Security Number (SCRSSN), (2) medical center identification number (3-digit numeric STA3N), (3) visit date (VIZDAY), and (4) clinic stop. Note that the clinic stop code variable is called CLNUM in the NPCD, which should have the same value as the variable called CL in the DSS file.

The results of the reconciliation were summarized in Table 18.

Table 18. Reconciliation between the NPCD Outpatient Event file and the DSS Outpatient Extract in FY2001 (Sample)

Records in DSS Outpatient Extract sample (NPCD only)	4,729,627
Records in NPCD Event file sample (excluding pharmacy consultation records)	5,079,023
Records in both files	4,727,933
Records only in the DSS Outpatient extract	1,694 (< 0.001%)
Records only in the NPCD Event file	351,090 (6.9%)

The reconciliation found that 7.0% of the records in the NPCD database did not have a corresponding DSS NPCD records (i.e., records with NPCD flag equal to "Y"). There are several explanations why these records were not included in the DSS Outpatient NPCD category. First, the DSS does not assign costs to telephone services. About 36% of the unmatched records in the NPCD Outpatient Event file were for various telephone services (Table 19). After excluding telephone-related encounters, there were still 4.4% of encounters in the NPCD Outpatient Event file that were not matched by the DSS records with NPCD flags.

Table 19. Telephone encounters in the unmatched NPCD Outpatient Event records FY2001
(Sample)

CL	CL Label	Frequency	Percent
324	Phone medicine	40111	11.42
147	Phone/ancillary	39849	11.35
527	phone general psych	15984	4.55
103	telephone triage	12092	3.44
424	phone surgery	5293	1.51
545	tele substance abuse	3380	0.96
546	telephone/MHICM	2643	0.75
216	phone rehab supp	1810	0.52
148	phone/diagnostic	1531	0.44
542	telephone PTSD	1298	0.37
528	phone/hmless ment ill	844	0.24
326	phone geriatrics	727	0.21
325	phone neurology	719	0.2
425	tele/prosth/orth	480	0.14
178	telephone/HBHC	432	0.12
536	tele/MH voc assist	193	0.05
729	telephone/domiciliary	167	0.05
565	telephone optometry	131	0.04
530	telephone/HUD-VASH	128	0.04
181	telephone/dental	126	0.04
537	tele psyc/soc rehab	91	0.03
579	tel/psychogeriatrics	50	0.01
606	telephone/chaplain	48	0.01
611	telephone dialysis	36	0.01
179	tele home care	23	0.01
	Total Telephone Services	128,186	36.51

Second, some outpatient services that were provided to patients who stayed in domiciliary or other residential facilities were included in the NPCD Outpatient Event file. The DSS system classified these services as inpatient care. We merged the unmatched records to the PTF extended bedsection file. If the visit date of an outpatient encounter was within the period of inpatient stay for the same patient, we identified those encounters as inpatient care. We found that 11,158 (3.2%) of the 351,090 unmatched NPCD Event records belonged to inpatient care. The frequency of these encounters and the associated inpatient bedsections were listed in Table 20.

Table 20. Unmatched outpatient encounters for patients stayed in long-term care facilities by bedsection FY2001

(Sample)

Bed No.	Bed Label	Frequency	Percent
85	Domiciliary	5804	52.02
25	Psyc Residential Rehab Trmt	1737	15.57
27	Sub Abuse Residential Rehab	1404	12.58
86	DOM Substance Abuse	911	8.16
80	Nursing home	429	3.84
26	PTSD Residential Rehab PRG	250	2.24
37	???	194	1.74
88	Dom PTSD	142	1.27
29	SA CWT/TR	104	0.93
28	HCMC CWT/TR	88	0.79
15	Gen (Acute) Med	28	0.25
74	Sub Abuse-Hi Int	21	0.19
91	Eval/BRF TRMT PTSD	19	0.17
20	Rehab Medicine	15	0.13
39	General CWT/TR	5	0.04
92	PSYC-General Intern	3	0.03
50	Surgery (GEN)	2	0.02
83	Respite Care	1	0.01
93	Hi Int Gen Psch-Inp	1	0.01
	Total	11,158	100.00%

The third explanation for the unmatched records is that the DSS and the NPCD use different methods to record an encounter for some services. For example, DSS uses date of blood draw (“accession”) as the encounter date, while NPCD may use the date test was run. Another example is when DSS obtains utilization from the radiology package and cannot find radiology care in NPCD, it assigns the encounter a stop code 105 (X-Ray). However, the utilization may enter the NPCD using a different stop code 149 (Radiation therapy treatment) or 151 (Magnetic resonance imaging).

The next analysis considered whether records that appeared to be unique in the NPCD file actually had corresponding care in the DSS file, but recorded with an incorrect flag, a different clinic stop, or a different data of care.

To examine these possibilities, we studied the encounters that appeared in the NPCD Outpatient Event file that did not have a corresponding DSS outpatient record with the NPCD flag, and did not involve telephone care. There were 222,904 of these unmatched NPCD outpatient encounters.

We considered whether the encounters were in the DSS file with an inappropriate flag variable. The DSS outpatient extract had 366,769 records for the sample patients with the flags for CLI and UTIL (records for clinical encounters).

We first considered if the DSS records had the same clinic stop, but a different visit date. We matched the two data sources using patient scramble social security number (SCRSSN), medical center identification number (STA3N) and clinic stop (CL in the DSS file and CLNUM in NPCD). If a person had multiple encounters in a clinic stop in both files, we considered the smaller number of encounters as matched. For example, if a person had three encounters in the X-Ray clinic (105) in the NPCD file and five X-Ray encounters in the DSS file, we considered the three records matched. Using this rule, we matched 45,055 (19%) outpatient encounters from the NPCD unmatched records; 60% of these records had the CLI in DSS and the remaining 40% the UTIL flag. Using this rule, there were 177,849 records from the sample patients that were in the NPCD but had no corresponding care in the DSS extract.

We then considered if the person had any other care recorded in the DSS file in the CLI and UTIL categories. We matched any encounters that belonged to the same person at that same medical center, regardless of the visit date or the clinic stop. This means that if a person had four unmatched outpatient encounters in the NPCD file and five encounters with the CLI or UTIL flag in the DSS file, four encounters were considered matched. Using the second rule, 136,528 (61%) of the 222,904 unmatched NPCD records were matched; 56% of the 136,528 matched records had the CLI flag in DSS and 44% the UTIL flag. Using this rule, there were 86,376 records from the sample patients that were in the NPCD but had no corresponding care in the DSS extract.

The first method is probably a reasonable upper limit for possible matches between the two databases. The second method of matching exhausted all possible matches. If a researcher is trying to link the DSS outpatient costs with the NPCD outpatient utilization, they should look at the records under the CLI and UTIL categories for possible matches.

Recommendations to researchers

Depending on the purpose of using the DSS outpatient extract, researchers should adopt different strategies. If researchers need to identify costs for outpatient pharmacy, the DSS outpatient pharmacy is probably the single source with such data. In the future, when the DSS pharmacy extract becomes available, researchers may compare it with the Pharmacy Benefit Management database. However, researchers should always examine cost outliers in the DSS outpatient pharmacy extract. Cost outliers should be validated and corrected, particularly for a study with small sample size.

For prosthetics, Denver distribution center, and Addiction Severity Index assessment, the DSS national extract is the only source for cost estimates. Again, researchers should examine the costs and may compare these costs with estimates from a micro-costing method. For encounters recorded in the NPCD outpatient file, researchers should first use the method described in this report to identify the unmatched encounters. Then, researchers could try to link the unmatched NPCD encounters with the records in the CLI and UTL categories in the DSS. The CLI category contains encounters recorded in the VistA, but not in the NPCD database. The UTL category contains utilization that cannot be linked to any encounter. As we have discussed in the report, researchers may need to drop the visit date or clinic stop name for possible matches. For the records that cannot be matched in any case, researchers can always impute their costs using the average costs of similar services in the DSS.

DSS treating specialty vs. DSS discharge files

Data in the DSS national data extracts discharge file were compared to the treating specialty file for FY01. The records in the treating specialty file were summarized so that they would have the same format as the discharge file, with one record per hospital stay. Stays that had not ended by the end of the fiscal year were excluded; although they are in the treating specialty file, they are not reported in the discharge file. Stays that began before 10/1/2000 were excluded because they are in the discharge file but not in the treating specialty file.

Table 21 shows how these sources reconciled. Although the two sources differ, they were in better agreement in FY01 than they were in FY00.

Table 21. Comparison of DSS national extract discharge and treating specialty files

Problem	FY2001	FY2000
Stays in the discharge file but not the treating specialty file	9	3,070
Stays in the treating specialty file but not in the discharge file	4,575	5,282
Stays in both files, but with costs that differed by more than \$100	153	3,292
Stays in both files, but with costs that differed by more than \$1,000	69	2,357

Stays Only in Treating Specialty File. Of the 4,575 stays in the treating specialty but not in the discharge file, a majority involved stays at five medical centers (see Table 22).

Table 22. The 5 medical centers with the most missing stays in the Discharge Extract

Station number	Number of stays in treating specialty file but not in discharge file
586	705
673	644
672	548
630	270
550	255
Total	2,422 (53%)

Differences in Cost between Files. There were 153 stays that had differences in cost of at least \$100 between treating specialty and discharge files. The cost was higher in the discharge file for 142 records; it was higher in the treating specialty file in 11 records.

There were a few records with noteworthy differences in cost, including one record in which the cost in the discharge file exceeded the cost in the treating specialty file by \$474,751. Two medical centers accounted for most of these problem records. Station 672 had 80% of the stays with cost differences that exceeded \$1,000. Station 672 had 43.8% and station 688 had 42.5% of the stays with cost differences that exceeded \$100.

Recommendations to researchers

Researchers should be aware of this possible discrepancy. When estimating costs for a study, the sum of bedsection costs may not equal to the sum of discharge costs if the data are obtained from the two DSS files separately.

Discussion

The results of the comparison of the DSS encounter-level extracts and the NPCD for FY2001 were very similar to the analysis conducted for FY2000. There was a good comparison between the DSS discharge file and the PTF main. The treating specialty file did not have a perfect match. The problem in matching these two sources of inpatient data appear to be largely due to differences in the way that they handle treatment provided in bedsections for less than 24 hours.

There continue to be substantial differences between DSS and the NPCD outpatient data in FY2001. A 10% sample of VA patients found that the DSS Outpatient Extract does not include any care for 1,758 people who appeared in the NPCD outpatient file. Further study on all patients from a single VISN showed the same percentage of missed people by the DSS Outpatient Extract. Most missed people had only one outpatient encounter during the year. Also excluded from the DSS extract are some 212,000 clinical encounters that appear in the NPCD, or 4% of the total. Some of these missing encounters may appear in the DSS file with an incorrect value of the flag variable that identifies the source of data, and with a different clinic stop or date of service from what is used by NPCD. Even under the most optimistic assumptions, however, at least half of the clinic encounters (or 2% of the total in the NPCD) cannot be found in the DSS data.

Compared with the results in FY2000, there was significant improvement in consistency between the DSS Treating Specialty Extract and the DSS Discharge Extract. Although hospital stays admitted and discharged within FY2001 should be recorded in both files, we still found that 4,575 (0.7%) stays in the Treating Specialty Extract were not matched in the Discharge Extract. We also found a few stays that had substantial difference in costs reported by the two files. Most of these inconsistencies, however, were concentrated in a few sites.

Reference

1. Yu W and P Barnett. Research Guide to Decision Support System National Cost Extracts. Menlo Park: VA HSR&D Health Economics Resource Center; 2002.