

Technical Report #9

**Reconciliation of DSS Encounter-Level National
Data Extracts and the VA National Patient Care
Database:
FY2001-FY2002**

**Wei Yu, PhD
Paul G Barnett**

December 2, 2003



*Dedicated to improving the quality
of health economics research*

Acknowledgements

Dr. Elisabeth McSherry and her colleagues at the DSS national program office provided valuable inputs for this report. We also thank Drs. Ann Hendricks and Douglas Bradham of HERC's expert panel for their useful comments. Ms. Ariel K. Hill provided excellent assistance in reviewing and editing this report.

Table of Contents

Executive Summary	iv
1. Introduction	1
2. Inpatient Discharges	2
2.1 Data	2
2.2 Methods	2
2.3 Results of reconciliation	3
2.4 Patterns of unmatched discharges	4
2.5 Summary of the discharge-file reconciliation	4
2.6 Recommendations to researchers	4
3. Inpatient Bedsection Stays	6
3.1 Data	6
3.2 Methods	6
3.3 Results of the treating-specialty reconciliation	8
3.4 Summary of treating-specialty reconciliation	10
3.5 Recommendations to researchers	10
4. Comparison between DSS treating specialty and discharge files	11
4.1 Stays only in treating specialty file	11
4.2 Differences in cost between files	11
4.3 Stays with negative cost	12
5. Outpatient encounters	13
5.1 Number of records and costs	13
5.2 Multiple categories	16
5.3 Cost outliers	16
5.4 Sample selection	21
5.5 Missed people	22
5.6 Multiple clinic stop encounters in a single visit	23
5.7 Results of outpatient reconciliation	23
5.8 Records only in NPCD	24
5.9 Records in other DSS categories	28
5.10 Recommendations to researchers	29
6. Summary	31
6.1 Bedsection admission and discharge dates	31
6.2 Missed outpatient records in DSS	31
6.3 Cost outliers	31
6.4 Value of DSS cost data to VA research	32
Reference	32
Appendix. Comparison of Outpatient Encounter Workload Between DSS & NPCD (Or Other Databases)	33

Tables

Table 2.1 Number of discharges in DSS NDE and NPCD PTF Inpatient discharge files FY2001 – FY2002.....	2
Table 2.2 Number of discharges in DSS NDE and PTF discharge files after excluding community nursing home stays and duplicates FY2001 – FY2002.....	3
Table 2.3 Reconciliation of DSS NDE discharge file with PTF Main files (FY2001).....	3
Table 2.4 Reconciliation of DSS NDE discharge file with PTF Main files (FY2002).....	3
Table 3.1 Records in DSS Treating Specialty and PTF Bedsection Files FY2001-FY2002.....	6
Table 3.2 DSS treating specialty file adjustments for the reconciliation FY2001 and FY2002.....	7
Table 3.3 Number of records in PTF bedsection files after deleting duplicates: FY2001 and FY2002.....	8
Table 3.4 Equivalent file names in PTF and DSS Treating Specialty File.....	8
Table 3.5. Reconciliation of treating specialty stays (FY2001).....	9
Table 3.6 Reconciliation of treating specialty stays (FY2002).....	9
Table 3.7 Differences in bedsection-in and out dates between DSS and PTF files among the unmatched records FY2002.....	10
Table 4.1 Comparison of DSS national extract discharge and treating specialty files.....	11
Table 4.2 Number of stays in treating specialty file not in discharge file.....	11
Table 5.1 DSS national outpatient extract FY2001-FY2002.....	13
Table 5.2 DSS outpatient clinic extract records classification.....	14
Table 5.3 Number of outpatient records and total costs for each data category FY2001 and FY2002.....	15
Table 5.4 Encounters and costs with two categories FY2002.....	16
Table 5.5 Cost outliers in the DSS outpatient extract (FY2001).....	17
Table 5.6 Stations with more than three cost outliers in the DSS outpatient clinic extract (FY2001).....	18
Table 5.7 Stations with more than three cost outliers in the DSS outpatient clinic extract (FY2002).....	18
Table 5.8 Stations with more than three cost outliers in the DSS outpatient pharmacy extract (FY2001).....	19
Table 5.9 Stations with more than three cost outliers in the DSS outpatient pharmacy extract (FY2002).....	19
Table 5.10 The top 10 clinic stops with outliers in the DSS outpatient clinic extract (FY2001).....	20
Table 5.11 The top 10 clinic stops with outliers in the DSS outpatient clinic extract (FY2002).....	20
Table 5.12 Stations with the top 10 cost outliers (FY2002).....	21
Table 5.13 A random sample of outpatient encounters in the DSS outpatient extract and NPCD outpatient SE file (FY2001 and FY2002).....	21
Table 5.14 Number of people who were recorded only in one database.....	23
Table 5.15 Reconciliation of outpatient records between the SE and the DSS NPCD category FY2001 and FY2002.....	24
Table 5.16 Telephone encounters in the unmatched NPCD outpatient event records FY2001.....	25
Table 5.17 Telephone encounters in the unmatched NPCD outpatient event records FY2002.....	26
Table 5.18. Outpatient encounters for inpatients by bedsection FY2001.....	27
Table 5.19. Outpatient encounters for inpatients by bedsection FY2002.....	27
Table 5.20 Other reasons for DSS non-matched outpatient records FY2002.....	29
Table 5.21 Costs of Pseudo SSNs in FY2002.....	29

Executive Summary

Linking demographic and clinical information in VA medical SAS files at Austin to cost information in the DSS National Data Extracts (NDEs) is useful for research and management. This report summarizes results of linking the DSS NDEs with the National Patient Care Database (NPCD) outpatient file and Patient Treatment files (PTF) in FY2002. Results of the FY2001 reconciliation are included for comparison. Detailed information on FY2001 data was reported in an earlier document¹.

Inpatient discharges The number of inpatient discharges recorded in the NDE discharge file and the three PTF discharge files were nearly identical. The number of mismatched discharges decreased further from more than 2000 discharges in FY2001 to less than 1000 discharges in FY2002 (less than 0.5%).

Inpatient bedsection stays VA characterizes hospital stays by segments based on bedsection, the type of care provided according to the treating specialty of the physician. The DSS treating specialty file and the NPCD bedsection files have different rules for setting bedsection admission and discharge dates. First, the DSS sets the admission date to the first day of the current fiscal year for any stays admitted in previous years. Second, the DSS does not count any day less than 24 hours whereas the NPCD does. About 23 percent of the total stays could not be matched because of these two design differences. After adjusting for those differences, about 5% of bedsection stays still could not be matched by the five variables: station number, patient ID, bedsection admission and discharge dates, and treating specialty.

DSS treating specialty vs. DSS discharge Stays and costs in the DSS treating specialty extract did not match those in the DSS discharge extract in FY2001 and FY2002 although the number of unmatched discharges and difference in cost were limited.

Outpatient encounters DSS identified more outpatient services than were recorded in the Austin NPCD. DSS allocated 22% of total outpatient costs to services that were not recorded in the NPCD in FY2002, such as prosthetics and addiction severity index tests. However, the DSS outpatient extract missed about 5% (3 million) of the records in the NPCD event file (the SE file), primarily due to delayed data entry in DSS. Any encounter coming to DSS after the previous month's DSS posting will not be assigned a cost. Starting from FY2003, the DSS NDE will have a separate file for outpatient encounters that are not assigned costs, which will improve the correspondence between the DSS and NPCD databases.

Outpatient cost outliers We identified any clinic encounter that cost \$100,000 or more as an outlier. In FY2002, there were 195 outliers (119 in FY2001) in the pharmacy file and 98 outliers (252 in FY2001) in the clinic file. Most (55%) of those non-pharmacy outliers involved prosthetics services.

Conclusion With adjustments for database design, inpatient stays in the DSS NDEs can be matched almost perfectly with corresponding records in the PTF. For outpatient services, the two databases differ largely in design. More than 90% of records in the NPCD event file can be linked to the DSS for cost information whereas DSS allocated 20% of outpatient cost to services other than those recorded in NPCD.

1. Introduction

This report presents results of reconciliation between the Decision Support System (DSS) National Data Extracts (NDEs) and files from the VA National Patient Care Database (NPCD) and Patient Treatment File (PTF) in fiscal year 2001 and 2002.

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 central office 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.

The VA National Patient Care Database (NPCD) and the Patient Treatment Files (PTF), also stored at the Austin computer center, include additional information about these encounters such as patient demographic characteristics, as well as the diagnoses and procedures associated with each encounter. These data are not included in the DSS national data 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 PTF and NPCD files. This report explains how the two databases can be matched. This reconciliation also has the value of examining whether the two databases completely characterize the services provided in the VA health care system.

HERC first reconciled the two databases for the federal fiscal year ending on September 30, 2000 (FY2000) and reported the results in a research guide on the use of the DSS extracts². We identified some discrepancies between the two databases, especially for outpatient care, in the FY2000 reconciliation. We also reconciled the FY2001 data with detailed analysis of outpatient data¹. That report updated the FY2000 reconciliation. In the following chapters, we report the reconciliation results for inpatient discharge, inpatient treating specialty, and outpatient files for FY2002.

2. Inpatient Discharges

Linking costs of each inpatient discharge in the DSS inpatient discharge file to diagnosis, treatment, and demographic information in the PTF discharge files is useful for healthcare studies. Although data included in the PTF are also available in the DSS production database, access to PTF is much easier than access to the DSS production database, especially for multi-site data. This chapter details how to link these two datasets and reports the matching results.

2.1 Data

The inpatient data are organized differently in the two databases (DSS and PTF). Discharges from all VA facilities including hospitals, nursing homes, residential programs, and domiciliary, are recorded in a single file in the DSS national data extract. Discharge records in the PTF database are divided between three 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. Table 2.1 lists the number of discharges reported in these three files for FY2001 and FY2002

**Table 2.1 Number of discharges in DSS NDE and NPCD PTF Inpatient discharge files
FY2001 – FY2002**

Data source	FY2001	FY2002
DSS NDE discharge file	689,820	630,955
PTF discharge files	705,290	690,180
PTF main acute discharge file (PM)	566,318	563,209
PTF main non-acute discharge file (XM)	79,285	77,043
PTF main observation discharge file (PMO)	59,687	49,928

Note: In FY2002, DSS classified discharges from observation beds as outpatient encounters. Therefore the total number of DSS discharges for FY2002 did not include discharges from observation bedsections.

2.2 Methods

Although the reconciliation method itself was straightforward, some adjustments were necessary before it could be completed.

Community nursing home stays

Before linking records from the two databases, we excluded discharges from community nursing homes that were recorded in the PTF XM file because DSS did not include discharges from community nursing homes. The community nursing home stays were identified by the variable STATYP=42 in the XM file. There were 11,141 discharges from community nursing homes in FY2001 and 7,862 in FY2002.

Merging variables

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).

Duplicates

Records with the same values for the four variables listed above were considered to be duplicates. In FY2001, there were 206 duplicates in the PM file, three duplicates in the XM file and 58 duplicates in the PMO file. In FY2002, there were 95 duplicates in the PM file, three duplicates in the XM file, and 10 duplicates in the PMO file. Overall, the number of duplicates dropped from FY2001 to FY2002. All duplicates were deleted before the merge.

Table 2.2 lists the number of records in each database after excluding community nursing home stays and duplicated records.

Table 2.2 Number of discharges in DSS NDE and PTF discharge files after excluding community nursing home stays and duplicates FY2001 – FY2002

Data source	FY2001	FY2002
DSS NDE discharge file	689,820	630,955
PTF discharge files	693,882	632,292
PTF main acute discharge file (PM)	566,112	563,114
PTF main non-acute discharge file (XM)	68,141	69,178
PTF main observation discharge file (PMO)	59,629	*

Note: Discharges from observation bedsections were excluded for the reconciliation in FY2002 because DSS classified those records as outpatient encounters.

2.3 Results of reconciliation

Tables 2.3 and 2.4 summarize the reconciliation of discharge files from FY2001 and FY2002, respectively.

Table 2.3 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 2.4 Reconciliation of DSS NDE discharge file with PTF Main files (FY2002)

	DSS NDE (n=630,955)	PTF Main (n=632,295)	
		PM (n=563,114)	XM (n=69,178)
Merge with all four variables	630,749 (99.96%)	562,127 (99.8%)	68,622 (99.2%)
Unmatched records	206	987	556

The FY2002 results were similar to those in FY2001. The two datasets reconciled almost perfectly (less than 1% of records were unable to be matched). The small differences may be due

to changes made after the PTF was closed in November and before the DSS NDE was generated the following Spring.

Some stays may have been excluded from DSS because the site was not current in processing DSS data. Also, some long-term care patients stayed for several years. The DSS could not estimate costs for patients admitted before DSS was implemented and those stays were not included in the DSS extracts. To identify these problems, unmatched records were further examined.

2.4 Patterns of unmatched discharges

First, we examined the 206 discharges only in DSS and found that 117 discharges matched with records in the observation bedsection file, suggesting that some sites did not correctly classify observation discharges as outpatient encounters. Next, we hypothesized that some of the unmatched discharges were due to errors or inconsistencies in recording values in the four merging variables, especially the admit and discharge dates. Therefore, we tried to rematch those unmatched discharges by dropping admit date, discharge date, or station number, respectively. After those extra steps, we were left with 75 discharges only in DSS, 981 discharges only in acute bedsections (PM), and 555 discharges in the extended bedsections (XM). Overall, those extra steps did not add a significant number of matches. We then further examined the patterns of those unmatched discharges.

Among the 75 DSS-only discharges, 12 were in Martinsburg and the others were distributed evenly among 39 other sites. Thirty six percent of these discharges were in general medicine and 16% in nursing homes. Most of them had missing values in discharge bedsection, suggesting some data quality problems with the unmatched records.

Among the 981 PTF-PM-only discharges, 248 (25%) occurred in a single station (Charleston), which suggested that these discharges were not included before the DSS NDE closing date. Further investigation showed that almost all the missed discharges in Charleston were in the Proctology bedsection.

For the 555 PTF-XM-only discharges, 68% were from nursing homes, 11% were from residential programs for substance abuse rehabilitation, and 12% were from VA domiciliary programs. While most of these discharges were evenly distributed among many stations, three stations (South Colorado, Martingsburg, and El Paso) accounted for 25% of the discharges that were not reported by DSS.

2.5 Summary of the discharge-file reconciliation

Compared to FY2001, the number of unmatched discharges in FY2002 declined.(less than 1%). Most of the unmatched discharges in the DSS database were likely due to missing values whereas most of the unmatched discharges in PTF may be due to missed records in the DSS national extracts (i.e., records not reported to NDEs before the closing date).

2.6 Recommendations to researchers

Researchers should be able to link the two datasets accurately for inpatient discharges. Although PTF contains a few more discharges than the NDE file, it is unlikely that the missed discharges

will affect the study results. Researchers can always check the PTF dataset to see if any extra discharges are missed. They can use the average costs of hospital stays with the same DRG to estimate costs of the discharges in PTF.

3. Inpatient Bedsection Stays

The DSS treating specialty file reports the monthly cost of each bedsection stay. It can be consolidated to give the total cost of each bedsection stay. This summary of the DSS treat specialty extract can be combined with the PTF bedsection files to create a file with cost and clinical information on each stay in a bedsection.

3.1 Data

Bedsection stays in the NDE treating specialty extract were reconciled with those in the PTF bedsection files. The DSS includes all records classified by inpatient treating specialty (i.e., bedsection) into a single file whereas the PTF separates bedsection records into three files: the main bedsection file (PB), the extended bedsection file (XB), and the observation bedsection file (PBO). In FY2002, DSS treated stays in observation bedsections as outpatient care. Therefore, records in observation bedsection were excluded from our FY2002 reconciliation. Table 3.1 documents the number of records for each file used for the reconciliation in the two datasets.

Table 3.1 Records in DSS Treating Specialty and PTF Bedsection Files FY2001-FY2002

Source	FY2001	FY2002
DSS NDE Treating Specialty file	1,293,596	1,211,258
PTF Bedsection files, total	907,460	877,880
PTF Acute bedsection file (PB)	765,099	747,840
PTF Non-acute bedsection file (XB)	82,620	80,097
PTF Observation bedsection file (PBO)	59,741	*

Note: The DSS NDE treating specialty file reports bedsection stay by fiscal period (month); the number of records in the treating specialty file is not equal to the number of bedsection stays (see details below).

Because the purpose of the DSS treating specialty file is to report the monthly cost of all inpatient stays, it reports the cost of a single bedsection stay in two or more records if the start and end dates of the stay across two or more months. For example, if a stay starts on January 20 and ends on February 5, the NDE treating specialty file would contain two records for the single stay; the first including the cost for the 11 days in January and the second for the four days in February. The PTF files, however, only include one record for each single stay. Also, the PTF has a census file that contains any stays that are not discharged or transferred by the end of fiscal year whereas the NDE treating specialty file includes those records in the last month of a fiscal year. Because of these structural differences between the two datasets, the number of records (not bedsection stays) in the NDE treating specialty file in Table 3.1 is larger than the total number of records (bedsection stays) in the three PTF files.

3.2 Methods

This section describes adjustments made for the reconciliation and the methods used in completing it. We use the term bedsection admission date for the date a patient is admitted or transferred to a bedsection and bedsection discharge date for the date a patient is discharged or transferred to another bedsection.

Census records

Prior to matching the records in the two databases, we made some adjustments. First, we dropped stays that were not discharged at the end of the fiscal year in the NDE Treating Specialty file (census stays). Although there is a PTF census file to match those census stays, this study does not attempt to reconcile it to the DSS treating specialty file. We excluded census stays in the NDE treating specialty extract by eliminating all records with a value of “Y” for the census stay variable (CENSUS).

Consolidation of DSS treating specialty records

Monthly records in the DSS treating specialty file were consolidated into one record for each unique bedsection stay. As discussed at the end of the previous section, 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 (month), 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 was consolidated into one record per bedsection stay using these variables.

Community nursing homes

Since the DSS NDE treating specialty extract did not contain data from community nursing homes, community nursing home stays from the PTF XB file were excluded by eliminating records with “STATYP=42.” Because the variable STATYP was kept in the main PTF extended care file, the main and bedsection files were merged to identify stays in community nursing homes.

Duplicated records

Records within a file with the same values in the five merging variables were considered to be duplicates and were removed before the merge.

Table 3.2 and 3.3 summarizes the adjustments described above for DSS TRT and PTF Bedsection files in FY2001 and FY2002, respectively.

Table 3.2 DSS treating specialty file adjustments for the reconciliation FY2001 and FY2002

Adjustments	FY2001 No. of Records	FY2002 No. of Records
Original	1,293,596	1,211,258
Non-discharged stays (census records)	156,728	152,261
Multiple records per bedsection stay	248,190	240,025
Records for reconciliation (unique bedsection stays)	888,678	818,972

Table 3.3 Number of records in PTF bedsection files after deleting duplicates: FY2001 and FY2002

Adjustments	Original	Community nursing homes	Duplicates	Records for reconciliation
FY2001 Total	907,460			895,118
Acute bedsection (PB)	765,099	-	1,078	764,021
Extended bedsection (XB)	82,620	11,145	28	71,447
Observation bedsection (PBO)	59,741	-	91	59,650
FY2002 Total	827,937			819,895
Acute bedsection (PB)	747,840	-	180	747,660
Extended bedsection (XB)	80,097	7,862	8	72,227

Note: observation stays were not included because they were not included in DSS in FY2002.

Merging variables

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 notice that three pairs of variables had different names in the two databases. The bedsection admission date was named as “BSINDAY” in the PTF and was named as “TRTIN” in the DSS treating specialty file. The bedsection discharge or transfer date was named as “BSOUTDAY” in the PTF and was named as “TRTOUT” in the DSS treating specialty file. The SAS name for bedsection number was BEDSECN in the PTF and was TRTSP in the DSS file. Table 3.4 lists the equivalent variables in the two datasets.

Table 3.4 Equivalent file names in PTF and DSS Treating Specialty File

PTF Names	Treating Specialty Names
BSINDAY	TRTIN
BSOUTDAY	TRTOUT
BEDSECN	TRTSP

3.3 Results of the treating-specialty reconciliation

Table 3.5 summarizes the reconciliation in FY2001. Please note that the observation bedsection stays were still included in the DSS treating specialty file in FY2001. For FY2002, we recognized a database design difference between DSS and PTF for stays in which patients were admitted before the first day of the current fiscal year and stayed into the current fiscal year. For those stays, the DSS treating specialty file sets the bedsection admission dates to the first day of the fiscal year (i.e., October 1, 2001 for FY2002). Another difference between the two files in database design was identified in FY2001. The DSS does not count a day that is fewer than 24 hours when a patient is transferred to another bedsection whereas the PTF does. Because of this difference, the DSS bedsection discharge dates are one day earlier than the PTF dates for stays in which patients are transferred to other bedsections. After adjusting for these two differences, we

found that 95% of the stays in DSS treating specialty file could be matched with corresponding stays in PTF (Table 3.6).

These two design differences have different impacts on the matching rates in the PB and XB files. Records in the PB file are bedsection stays in an acute hospital. Table 11 summarized these patterns. Because patients who stay in an acute hospital are more likely to be transferred to different bedsections, 83% of the unmatched stays in the PB file are due to the one-day difference in bedsection discharge dates (Table 3.7). The XB file contains stays in nursing homes, domiciliary facilities, and residential programs. Stays in the XB have much longer length of stay than stays in the PB file. Therefore, more stays in the XB file are admitted in the previous years. We found that 75% of the unmatched extended bedsection stays are due to the truncation in bedsection admission dates. Although about 5% of the bedsection stays are still not matched after adjustments for the design differences, the numbers of unmatched stays in the two datasets are close. With further examination and adjustments, most of the unmatched stays are likely to be linked.

Table 3.5. Reconciliation of treating specialty stays (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%)

Table 3.6 Reconciliation of treating specialty stays (FY2002)

	DSS NDE (n=818,972)	PTF Bedsection Files (n=819,895)	
		PB (n=747,660)	XB (n=72,227)
Merge with all five variables	594,310	538,951	55,359
Non-matched records (percent of the total)	224,604 (27.4%)	208,679 (27.9%)	16,862 (30.5%)
Unmatched number of records after adjustments for inconsistency in bedsection admission and discharge dates (percent of the total)	27,999 (4.7%)	25,631 (4.8%)	3,305 (6.0%)

**PBO records were not reconciled in FY2002 because DSS classified observation stays as outpatient care.*

Table 3.7 Differences in bedsection-in and out dates between DSS and PTF files among the unmatched records FY2002

	DSS TRT	PTF PB	PTF XB
Unmatched number of record	224,604	208,679	16,862
Matched with adjustment for DSS bedsection discharge date being one day earlier than PTF (percent of the total unmatched stays)	174,126 (78%)	172,450 (83%)	1,676 (10%)
Matched by setting BSINDAY as 10/1/2001 for stays admitted before 10/1/2001 (percent of the total unmatched stays)	22,479 (10%)	10,598 (5%)	11,881 (71%)
Unmatched number of records after adjustments for inconsistency in bedsection admission and discharge dates (percent of the total unmatched records)	27,999 (13%)	25,631 (8%)	3,305 (20%)

3.4 Summary of treating-specialty reconciliation

After adjusting for the design difference in bedsection admission and discharge dates, we were unable to match 5% of total stays using the five merging variables. The significant increase in the matching rate from FY2001 to FY2002 is mainly due to the recognition of different ways of recording bedsection admission dates for stays in which patients were admitted in previous years.

3.5 Recommendations to researchers

If researchers need to link bedsection stays in the DSS and PTF datasets, they should first extract all bedsection stays from the NPCD and DSS databases for the study subjects using scrambled social security number regardless of other information. Then, researchers can adopt the method used in this section to conduct a match. For the 5% unmatched stays, researchers may link them in the two datasets by dropping STA3N, admission and discharge dates. For the very few unmatched records, manual examination is necessary.

4. Comparison between DSS treating specialty and discharge files

To validate whether stays and costs reported in the DSS treating specialty file are consistent with stays reported in the DSS discharge file, we reconciled these two files for FY2002. 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 the first day of the fiscal year (10/1/2001 for FY2002) were excluded because costs were included in the discharge file but not in the treating specialty file.

Table 4.1 shows how these sources reconciled. Discrepancies increased over levels observed in FY2001 and FY2002.

Table 4.1 Comparison of DSS national extract discharge and treating specialty files

Problem	FY01	FY02
Stays in the discharge file but not the treating specialty file	9	1,050
Stays in the treating specialty file but not in the discharge file	4,575	5,663
Stays in both files, but with costs that differed by more than \$100	153	203
Stays in both files, but with costs that differed by more than \$1,000	69	136

4.1 Stays only in treating specialty file

Of the 5,663 stays in the treating specialty file (not in the discharge file) a majority involved stays at only a few medical centers (see Table 4.2).

Table 4.2 Number of stays in treating specialty file not in discharge file

Top 5 stations with problems in FY2002

Station number		FY01	FY02
586	Jackson, MS	705	791
672	San Juan, PR	548	542
664	San Diego, CA	132	416
630	New York, NY	270	339
540	Clarksburg, WV	20	226

Stations 664 and 540 were newly added to the list of top 5 problem stations in FY2002; they previously ranked #6 and #38. Stations 673 and 550, which were among the top 5 with problems in FY01, had far fewer mismatches in FY02.

4.2 Differences in cost between files

There were 203 stays that had differences in cost of at least \$100 between the treating specialty and discharge files. The cost was higher in the discharge file for 160 records; it was higher in

the treating specialty file for 43 records. Unlike previous years, these problems were widely spread among different medical centers.

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 \$173,800.

4.3 Stays with negative cost

There were 15 stays that were assigned a negative cost. These stays were assigned identical negative costs in both files. One of these stays was assigned a cost of -\$53,116.

5. Outpatient encounters

The DSS outpatient extract contains services recorded in the NPCD. The NPCD outpatient event file (the SE file) includes all encounters to outpatient clinic stops. Linking records in the NPCD SE file with the DSS outpatient extract can generate a combined outpatient file with cost and clinical information, which is useful for VA healthcare studies. The DSS outpatient file is designed to include many outpatient services that are not recorded in the NPCD database. We first identified which records in the DSS outpatient extract used NPCD as their data source. Then we matched records in DSS that were marked by the NPCD flag with records in the NPCD SE file, using a random sample of 10 percent of the VA patients in FY2002.

5.1 Number of records and costs

The DSS Outpatient extract consists of two data sets: the outpatient pharmacy data and the outpatient clinic data. Each dataset includes four files separated by VISNs: VISN 1 – VISN 6, VISN 7 – VISN 11, VISN 12 – VISN16, and VISN 17 – VISN 22. The outpatient pharmacy extract contains VA pharmacy utilization and the outpatient clinic extract includes all other outpatient encounters. The number of records and total costs are listed in Table 5.1.

Table 5.1 DSS national outpatient extract FY2001-FY2002

File	FY2001	FY2002
Number of records		
Outpatient Clinic	62,914,917	65,232,522
Outpatient Pharmacy	47,961,983	53,462,803
Costs		
Outpatient Clinic	\$9,157,119,228	\$9,994,361,606
Outpatient Pharmacy	\$3,061,169,153	\$3,498,881,678

DSS allocated costs to outpatient activities recorded in the VISTA system. Because some activities at VISTA were not recorded in the NPCD database, DSS allocated costs to more activities than the workload included in the NPCD database. To identify those activities, the DSS outpatient extract included eight categories classified by seven flag variables, each representing a primary data source, and an eighth category, without a flag variable. Table 5.2 describes the eight categories and Table 5.3 summarizes the number of records and costs for each category in the FY2001 and FY2002 clinic outpatient extract.

Table 5.2 DSS outpatient clinic extract records classification

Flag Variable	Description
PRE	Outpatient pharmacy clinic utilization 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.
All Other (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 5.3 Number of outpatient records and total costs for each data category FY2001 and FY2002

Category	Records			
	FY2001		FY2002	
	N	%	N	%
NPCD	50,771,919	80.7	52,605,486	80.6
CLI	1,425,256	2.3	1,321,373	2.0
PROS	2,136,597	3.4	2,260,169	3.5
DDC	461,320	0.7	492,175	0.8
NOSHOW	4,522,878	7.2	4,375,320	6.7
PRE	0	0	10,052	0.0
UTIL	2,485,305	4.0	2,402,130	3.7
All Other*	930,777	1.5	1,514,062	2.3
MULTIPLE	180,865	0.3	272,486	0.4
Total	62,914,917	100.0	65,253,253	100.0
Category	Cost			
	FY2001		FY2002	
	\$	%	\$	%
NPCD	7,258,361,209	79.3	7,826,490,877	78.27
CLI	188,534,076	2.1	205,403,391	2.05
PROS	457,782,594	5.0	505,281,859	5.05
DDC	92,021,366	1.0	81,984,954	0.82
NOSHOW	222,722,213	2.4	238,375,954	2.38
PRE	0	0	12,178,356	0.12
UTIL	380,680,774	4.2	327,356,465	3.27
All Other*	516,045,931	5.6	730,524,800	7.31
MULTIPLE	40,971,065	0.4	72,077,824	0.72
Total	9,157,119,228	100.0	9,999,674,480	100.00

*"All other" is a group identified by excluding records marked by the seven flags.

For both FY2001 and FY2002, the services recorded in the NPCD database accounted for only 80% of the total DSS outpatient cost. However, the FY2002 data had a couple of new problems. First, 10,052 pharmacy records were included in the outpatient clinic file for a total cost of \$12 million. It is unclear whether these records were different from those in the pharmacy file or were included in the clinic file by mistake. Second, the number of records simultaneously classified into two categories increased. The problem of multi-category records is further investigated and reported in the following section.

5.2 Multiple categories

According to the National DSS Extract Technical Guidebook (DSS BTSSO/Development, April 3, 2000), each encounter should be classified into a single category with the highest priority for NPCD records. However, about 0.4% of outpatient records were assigned to two categories. Compared with FY2001, the number of multi-category records increased from 180,865 to 272,486 and costs increased from \$41 million to \$72 million in FY2002.

Further analysis showed that more than 90% of these multiple-category records were in two groups. The largest group combined NPCD and PROS (prosthetics), accounting for 71% of the multiple-category records and 56% of the cost (Table 5.4). The second largest group was a combination of NPCD and UTIL, accounting for 24% of multi-category records and 22% of the cost. It was unclear why the NPCD category flag did not replace the PROS and UTIL flags.

Table 5.4 Encounters and costs with two categories FY2002

Combination	SAS Value*	No. of records	%	Costs	%
NPCD + PROS	YNYNNNN	192,473	70.6	40,422,557	56.1
NPCD + UTIL	YNNNNNY	63,988	23.5	15,783,538	21.9
CLI + PROS	NYYNINN	10,238	3.8	4,531,538	6.3
CLI + PRE	NYNNNYN	5,235	1.9	10,075,250	14.0
Others	Others	552	0.2	1,264,942	1.8
Total		272,486	100.0	72,077,824	100.0

* The value of SAS variable ENCFLAG. It reflects the eight possible categories to which an encounter can be classified. For example, an encounter in the NPCD group should have the first letter equal to "Y" and other letters equal to "N".

5.3 Cost outliers

Records with a total cost of \$100,000 or higher for a single clinical encounter were identified from both the DSS outpatient clinic and pharmacy extracts. We tabulated those outliers in Table 5.5. Compared with FY2001, the FY2002 data had fewer outliers in the clinic file and more outliers in the pharmacy file. The largest outlier in a single pharmacy utilization was assigned a cost of \$1.25 million.

Table 5.5 Cost outliers in the DSS outpatient extract (FY2001)

(Total cost >= \$100,000/record)

	FY2001	FY2002
Pharmacy		
Number of records	119	195
Total cost	\$27,290,686	\$43,034,998
Maximum cost /record	\$860,167	\$1,249,454
Other Clinics		
Number of records	152	98
Total cost	\$31,308,888	\$22,180,695
Maximum cost / single stop visit	\$866,797	\$758,844

We further examined these cost outliers by medical center (STA3N) and clinic stop (for clinic outliers only) and reported the results in Tables 5.6 through 5.12. In the outpatient clinic file, the number of stations with more than three outliers dropped from 17 in FY2001 to 9 in FY2002 (Tables 5.6 and 5.7). In the pharmacy file, however, the number of stations with three or more outliers increased from 9 in FY2001 to 12 in FY2002 (Tables 5.8 and 5.9). For a few stations, cost outliers occurred in both years (e.g., New York Harbor, Baltimore, Boston).

Table 5.10 and 5.11 lists cost outliers in outpatient clinic and pharmacy files, respectively. The data indicates that about half of the cost outliers in the clinic file are from prosthetics. Although some items in the prosthetics category are expensive, further analysis is still needed. Cost outliers in the pharmacy file are clearly problematic. Table 5.12 lists stations that had the top 10 most expensive cost outliers. Station 619 (Montgomery) has five of the top 10 most expensive outliers, all with the same value (\$593,784) in the clinic file and Station 512 (Baltimore) has 7 of the top 10 outliers, all with very similar values (\$623,446 - \$623,537).

Table 5.6 Stations with more than three cost outliers in the DSS outpatient clinic extract (FY2001)

(Total cost >=\$100,000/record)

STA3N	STA3N Label	Frequency	Percent
618	Minneapolis	15	9.87
541	Cleveland-Wade Park	14	9.21
520	Gulf Coast HCS	11	7.24
521	Birmingham	10	6.58
580	Houston	10	6.58
630	N.Y. Harbor HCS	9	5.92
619	Montgomery	7	4.61
512	Baltimore	6	3.95
671	San Antonio	6	3.95
459	Honolulu	5	3.29
629	New Orleans	5	3.29
528	Upstate N.Y.	4	2.63
554	Denver	4	2.63
578	Hines	4	2.63
660	Salt Lake City Healthcare	4	2.63
537	VA Chicago HCS	3	1.97
598	Little Rock	3	1.97

Table 5.7 Stations with more than three cost outliers in the DSS outpatient clinic extract (FY2002)

(Total cost >=\$100,000/record)

STA3N	STA3N Label	Frequency	Percent
523	Boston	23	23.47
521	Birmingham	10	10.2
619	Montgomery	7	7.14
660	Salt Lake City Healthcare	6	6.12
673	Tampa	6	6.12
554	Denver	4	4.08
691	La Wadsworth	4	4.08
442	Cheyenne	3	3.06
546	Miami	3	3.06

Table 5.8 Stations with more than three cost outliers in the DSS outpatient pharmacy extract (FY2001)

(Total cost >=\$100,000/record)

STA3N	STA3N Label	Frequency	Percent
630	N.Y. Harbor	52	43.7
405	White River Jct	11	9.24
452	VAMC Wichita KS<102001	10	8.4
528	Upstate N.Y.	8	6.72
553	Detroit VAMC	7	5.88
642	Philadelphia	4	3.36
629	New Orleans	3	2.52
657	VA Heartland-E VH MO	3	2.52
662	San Francisco	3	2.52

Table 5.9 Stations with more than three cost outliers in the DSS outpatient pharmacy extract (FY2002)

(Total cost >=\$100,000/record)

STA3N	STA3N Label	Frequency	Percent
630	N.Y. Harbor	66	33.85
660	Salt Lake City	28	14.36
512	Baltimore	15	7.69
523	Boston	14	7.18
632	Northport	13	6.67
655	Saginaw	9	4.62
673	Tampa	9	4.62
463	Anchorage	8	4.1
637	Asheville-Oteen	5	2.56
437	Fargo	4	2.05
570	Fresno	4	2.05
657	St. Louis-John Cochran	4	2.05

Table 5.10 The top 10 clinic stops with outliers in the DSS outpatient clinic extract (FY2001)

(Total cost >=\$100,000/record)

CL	CL Label	Frequency	Percent
423	Prosthetics	67	44.08
146	PET	14	9.21
125	Social work services	7	4.61
429	Outpatient care in the O.R. room	7	4.61
553	Day treatment group	7	4.61
UNK		5	3.29
323	Primary Care/Med	5	3.29
650	Unknown	4	2.63
105	X-Ray	3	1.97
108	Laboratory	3	1.97
Xxx	All Other clinic stops	30	19.7

Table 5.11 The top 10 clinic stops with outliers in the DSS outpatient clinic extract (FY2002)

(Total cost >=\$100,000/record)

CL	CL Label	Frequency	Percent
423	Prosthetics	54	55.1
681	Contracted nursing homes	11	11.22
117	Nursing	5	5.1
105	X-Ray	4	4.08
203	Physical Therapy	3	3.06
291	Observation Surgery	3	3.06
421	Vascular Lab	3	3.06
429	Outpatient Care in O.R.	3	3.06
108	Laboratory	2	2.04
102	Admit/Screening	1	1.02
Xxx	All Others	9	9.0

Table 5.12 Stations with the top 10 cost outliers (FY2002)

STA3N	STA3N Label	Clinic cost outliers (\$)	STA3N	STA3N Label	Pharmacy cost outliers (\$)
561	East Orange	758,844	632	Northport	1,249,454
612	NCHC Martinez	683,460	660	Salt Lake City HCS	655,732
534	Charleston	606,939	512	Baltimore	623,738
619	Montgomery	593,784	512	Baltimore	623,537
619	“	593,784	512	Baltimore	623,509
619	“	593,784	512	Baltimore	623,487
619	“	593,784	512	Baltimore	623,464
619	“	593,784	512	Baltimore	623,462
541	Cleveland Wade-Park	591,989	512	Baltimore	623,446
636	VA HB Western IA HCS	540,934	512	Baltimore	623,445

5.4 Sample selection

Because the number of outpatient records was very large, we selected a random sample of VA patients for the reconciliation (about 10% of total patients). In FY2002, we selected a sample using a different range of scrambled SSNs from those selected in FY2001. We extracted all the outpatient records from the two databases for people whose last two digits of the scrambled Social Security Numbers were between 54 and 57 (including 54 and 57) or between 64 and 67 (including 64 and 67).

Table 5.13 compares data on the sample cohort from the two sources. Please note that the number of records and people are not comparable between the two years because we selected a different group of people in FY2002 from FY2001.

Table 5.13 A random sample of outpatient encounters in the DSS outpatient extract and NPCD outpatient SE file (FY2001 and FY2002)

Data Source	FY2001	FY2002
Number of People		
NPCD Outpatient SE file	310,102	361,453
DSS Outpatient Clinic Extract (NPCD flag)	306,057	357,712
Number of Records		
NPCD Outpatient SE file	5,653,059	5,281,173
DSS Outpatient Clinic Extract (NPCD flag)	4,729,627	4,349,960

Table 5.13 shows that the differences between the DSS NPCD records and the records in the Austin NPCD SE file are consistent across the two years. In FY2001 and FY2002, the number of records in the NPCD outpatient SE file are larger than the number of NPCD records in the DSS outpatient clinic extract; the number of people are also larger in the SE file than the number of people included in the DSS NPCD records.

5.5 Missed people

We also examined the number of people who were only recorded in one database. Table 5.14 summarizes these findings for FY2001 and FY2002.

In FY2001, there were 3,203 people who only appeared in the NPCD SE file and 163 people who only appeared in DSS outpatient clinic file with a visit marked by the NPCD flag. In FY2002, there were 3,849 people who only appeared in the NPCD SE file and 190 people who only appeared in the DSS file with an NPCD flag.

For FY2001, we further examined whether these 3,203 people had any records in other categories of the DSS outpatient file. We found that 1,445 people had some records in other DSS outpatient categories, but 1,758 people in the NPCD SE file had no records in any DSS outpatient categories. If the selected sample was representative of the entire VA population, the estimated number of people who had no record in any DSS outpatient categories would be 19,533 in FY2001. Results from a study in VISN 21 data indicated that most of those missed people had only one encounter during the entire fiscal year which explains why some centers did not report the workload in time. Patients names were eliminated from the FY2002 NDE database if their single encounter were not reported in time to DSS.

We did similar analysis for FY2002. We found that among the 3,849 people who appeared in the NPCD SE file, but had no records in DSS with an NPCD flag, 1,720 had some records in DSS with other flags. However, 2,139 people in the NPCD SE file had no records in any category of the DSS outpatient file. Using this rate of missing people derived from the sample, we estimated that 26,738 people may be left without any records in the FY2002 DSS outpatient file. Researchers should understand that DSS included outpatient services that were not recorded in the NPCD SE file. For fiscal 2002, 18,618 people had records in DSS with only non-NPCD flags, and 60% of those people only had records with the “No Show” flag.

Table 5.14 Number of people who were recorded only in one database

(Based on a random sample of outpatient encounters in the DSS outpatient extract and NPCD outpatient SE file in FY2001 and FY2002)

Data Source	FY2001	FY2002
Patients only in the NPCD SE file		
Number of patients in the NPCD SE file, but not in the DSS records with NPCD flag (Expected number of missing people in the entire population)*	3,203 (35,589)	3,849 (48,113)
Number of patients in the NPCD SE file, but not in any DSS records (Expected number of missing people in the entire population)*	1,758 (19,533)	2,139 (26,738)
Patients only in the DSS outpatient file		
Patients only in the DSS outpatient extract with NPCD flag (Expected number of missing people in the entire population)*	163 (1,811)	190 (2,375)
Patients only in the DSS outpatient extract for services that are not in the NPCD.	NA	18,618

**The expected number for the entire population was calculated based on the percent of records in each random sample (9% for FY2001 and 8% for FY2002).*

5.6 Multiple clinic stop encounters in a single visit

A patient can have multiple NPCD records with the same clinic stop during a single day. The DSS outpatient file, however, consolidates multiple encounters in a single day to a single clinic stop in one record for each day. To reconcile the two datasets, therefore, we consolidated these multiple encounters in the NPCD SE file. In FY2002, 440,453 (8.4%) of the 5,236,396 records in the NPCD SE sample were additional encounters to a single clinic stop during a single day. We checked whether the DSS outpatient file contained any duplicated records - records with the same values for person ID (SCRSSN), station (STA3N), clinic stop (CL), and visit day (VIZDAY). We found no duplicates in the records with NPCD flags, but 37 duplicates (of 975,407) among the records with other flags.

5.7 Results of outpatient reconciliation

As described in Table 5.3, the DSS outpatient extract contains many more services than are recorded in the NPCD SE file. Records with NPCD flags and those in the NPCD SE file were reconciled in DSS. After consolidating multiple clinic encounters incurred by a patient during a single visit in the NPCD SE file, 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. Researchers should notice that the SAS variable name for clinic stop was CLNUM in the NPCD and CL in the DSS file. Table 20 summarizes the reconciliation.

Table 5.15 Reconciliation of outpatient records between the SE and the DSS NPCD category FY2001 and FY2002

	FY2001 sample	FY2002 sample
Records in DSS Outpatient Extract sample with NPCD flag	4,729,627	4,349,960
Records in consolidated NPCD SE file sample (excluding pharmacy consultation records)	5,079,023	4,795,943
Records in both files	4,727,933	4,347,726
Records only in the DSS Outpatient extract	1,694 (< 0.001%)	2,234 (<0.001%)
Records only in the NPCD SE file	351,090 (6.9%)	448,217 (9.3%)

Note: The FY2001 sample was slightly larger than the FY2002 sample. Therefore, the numbers between the two years are not comparable. Only the percentages have comparative meaning.

We found that in FY2001, 6.9% of the records in the NPCD SE file did not have a corresponding DSS NPCD record (i.e., records with NPCD flag equal to “Y”). The percentage of non-matched records increased to 9.3% in FY2002.

5.8 Records only in NPCD

The reconciliation results showed that almost all records with NPCD flag in the DSS outpatient file had corresponding records in the NPCD SE file. Approximately 5% of the records in the consolidated NPCD SE file were left without corresponding records in the DSS. The DSS Bedford Technical Center provided a list of possible reasons that some records in the two databases could not be matched (Appendix). We explored three major reasons for the NPCD-only records: telephone services, services for inpatients, and delayed data entry.

Telephone services

DSS did not assign costs to telephone services. In FY2001, 36% of the unmatched records in the NPCD outpatient SE file were for various telephone services (Table 5.16). In FY2002, telephone services accounted for 34% of the unmatched clinic encounters (Table 5.17). After excluding telephone-related encounters, 4.4% of encounters in the consolidated NPCD outpatient SE file in FY2001 could not be matched to the DSS outpatient NDE. This increased to 6% in FY2002.

Table 5.16 Telephone encounters in the unmatched NPCD outpatient event records FY2001
 (Based on the 10 percent sample of patients)

CL	CL Label	Frequency
324	PHONE MEDICINE	40111
147	PHONE/ANCILLARY	39849
527	PHONE GENERAL PSYCH	15984
103	TELEPHONE TRIAGE	12092
424	PHONE SURGERY	5293
545	TELE SUBSTANCE ABUSE	3380
546	TELEPHONE/MHICM	2643
216	PHONE REHAB SUPP	1810
148	PHONE/DIAGNOSTIC	1531
542	TELEPHONE PTSD	1298
528	PHONE/HMLESS MENT ILL	844
326	PHONE GERIATRICS	727
325	PHONE NEUROLOGY	719
425	TELE/PROSTH/ORTH	480
178	TELEPHONE/HBHC	432
536	TELE/MH VOC ASSIST	193
729	TELEPHONE/DOMICILIARY	167
565	TELEPHONE OPTOMETRY	131
530	TELEPHONE/HUD-VASH	128
181	TELEPHONE/DENTAL	126
537	TELE PSYC/SOC REHAB	91
579	TEL/PSYCHOGERIATRICS	50
606	TELEPHONE/CHAPLAIN	48
611	TELEPHONE DIALYSIS	36
179	TELE HOME CARE	23
	Total Telephone Services	128,186
	Percent of records only in NPCD not found in DSS	36.51%

Table 5.17 Telephone encounters in the unmatched NPCD outpatient event records FY2002
(Based on the 10 percent sample of patients)

CL	CL Label	Frequency
147	PHONE/ANCILLARY	55599
324	PHONE MEDICINE	52756
103	TELEPHONE TRIAGE	15309
527	PHONE GENERAL PSYCH	10371
424	PHONE SURGERY	6072
216	PHONE REHAB SUPP	2861
545	TELE SUBSTANCE ABUSE	1204
425	TELE/PROSTH/ORTH	1180
178	TELEPHONE/HBHC	979
326	PHONE GERIATRICS	675
542	TELEPHONE PTSD	674
546	TELEPHONE/MHICM	588
325	PHONE NEUROLOGY	403
148	PHONE/DIAGNOSTIC	384
528	PHONE/HMLESS MENT ILL	369
428	TELEPHONE OPTOMETRY	164
579	TEL/PSYCHOGERIATRICS	143
181	TELEPHONE/DENTAL	113
530	TELEPHONE/HUD-VASH	104
729	TELEPHONE/DOMICILIARY	73
536	TELE/MH VOC ASSIST	63
537	TELE PSYC/SOC REHAB	28
611	TELEPHONE DIALYSIS	23
179	HOME TELEVIDEO CARE	13
169	TELEPHONE/CHAPLAIN	12
	Total Telephone Services	150,160
	Percent of records only in NPCD not found in DSS	33.50%

Services for inpatients

Some outpatient services that were provided to patients who stayed in domiciliary or other residential facilities were included in the NPCD SE 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 the encounters as inpatient care. We found that 11,158 (3.2%) of the 351,090 unmatched NPCD Event records corresponded to an inpatient stay in FY2001 (Table 5.18). The frequency of this problem dropped substantially to 484 (1.1%) in FY2002 (Table 5.19).

Table 5.18. Outpatient encounters for inpatients by bedsection FY2001

(Based on the 10 percent sample of patients)

Bedsection	Bed Label	Frequency
85	DOMICILARY	5804
25	PSYC RESIDENTIAL REHAB TRMT	1737
27	SUB ABUSE RESIDENTIAL REHAB	1404
86	DOM SUBSTANCE ABUSE	911
80	NURSING HOME	429
26	PTSD RESIDENTIAL REHAB PRG	250
37	DOMICILARY CARE FOR HOMELESS VETERANS	194
88	DOM PTSD	142
29	SA CWT/TR	104
28	HCFI CWT/TR	88
15	GEN (ACUTE) MED	28
74	SUB ABUSE-HI INT	21
91	EVAL/BRF TRMT PTSD	19
20	REHAB MEDICINE	15
39	GENERAL CWT/TR	5
92	PSYC-GENERAL INTERN	3
50	SURGERY (GEN)	2
83	RESPITE CARE	1
93	HI INT GEN PSCH-INP	1
	Total	11,158
	Percent of records only in NPCD not found in DSS	2.5%

Table 5.19. Outpatient encounters for inpatients by bedsection FY2002

(Based on the 10 percent sample of patients)

Bedsection	Bed Label	Frequency
80	NURSING HOME CARE	166
85	DOMICILIARY	155
86	DOM SUBSTANCE ABUSE	46
37	DOM CARE HMLS(DCHV)	31
27	SUB ABUSE RES REHAB	23
88	DOM PTSD	14
25	PSYC RES REHAB TRMT	12
26	PTSD RES REHAB PGM	10
29	SA CWT/TR	9
28	HCFI CWT/TR	6
15	GEN(ACUTE) MED	4
40	INTERMEDIATE MED	3
81	GEM NHC	2
20	PTSD/CWT/TR	1
38	REHAB MEDICINE	1
83	RESPITE CARE	1
	Total	484
	Percent of records only in NPCD not found in DSS	0.1%

Later data entry

Telephone services and services provided to inpatients only explained 34% of the unmatched records. The rest of the records that were only in the NPCD may have been due to delayed transmission of data to DSS. DSS NDEs only include those services that have allocated costs. If a site entered services provided in the previous month after the previous month's DSS posting, those services would not receive costs. Also DSS stops making changes and collecting data on October 19 of each year whereas the Austin NPCD database continues to accept changes and new data after that date. Starting from FY2003, the DSS NDE will include services that are not assigned costs in a separate file. We will then be able to learn if most of the unmatched records in the NPCD SE files are matched by no-cost records in DSS.

5.9 Records in other DSS categories

Based on the explanations for unmatched records provided by the BTSO memo, we also examined four groups of data that would be useful for researchers: observation bedsection records, radiology, non-VA encounters, and pseudo SSNs. Because most of the data examined under the four groups were not under the NPCD flag, we examined these data in DSS records with other flags as well as the 2,234 records with NPCD flag that were left without a corresponding record in the NPCD SE file. To avoid confusion, we call them as DSS-only records.

Observation bed encounters

According to the BTSO Memo, observation encounters were reported in DSS as outpatient encounters using stop codes 290-296. The Austin database still recorded observation encounters as inpatient stays. In FY2002, there were 4,093 records in the observation group, accounting for less than 1% of the DSS-only records (Table 5.20).

Radiology

Radiology encounters were assigned several different clinic numbers in Austin NPCD SE file but were assigned to stop 105 in DSS. According to the BTSO Memo, this issue was resolved in FY2002. However, we still observed 13,166 DSS non-matched records in stop code 105, accounting for more than 1% of the DSS-only records (Table 5.20).

Non-VA encounters

Non-VA care was recorded in DSS as stop codes 650-658 and 681-685 for contract nursing homes. In FY2002, there were 34,575 records belonging to these stop codes, accounting for 3.5% of the DSS-only records (Table 5.20). These records include contract nursing home stays and visits by an RN to patients in contract nursing homes. Please notice that for FY2002, the DSS NDE did not have information about the length of stay for these community nursing home encounters. Therefore, costs allocated to such encounters might not be accurate because length of stay was unknown. The data in FY2002 indicated that these services cost \$39.4 million. The median cost was \$546, the maximum cost was \$107,323, and 25% of those encounters cost \$74 or less that may be associated with the RN visits. There was one encounter with a negative cost. Some of the non-VA nursing home utilization were visits by a VA nurse to the nursing home (stop 680) which is why costs may seem low.

Table 5.20 Other reasons for DSS non-matched outpatient records FY2002

Category	Frequency	Percent of DSS-only records*
Contract nursing homes	34,575	3.5
Radiology	13,166	1.3
Observation bed	4,093	0.4

* Include the 2,234 records that were under the NPCD flag and only in DSS as well as DSS records under all other flags.

Pseudo SSNs

Some records in the DSS system could not be linked to an individual, such as a reference test in laboratory. When a laboratory conducted a reference test, this service and its associated cost was distributed to a pseudo SSN. Because of the large volume of reference tests, the total cost for this pseudo SSN could be significantly large. The pseudo SSN was not excluded from the DSS NDEs. There were other situations where services were not linked to a person and a pseudo SSN was used. In these situations a common feature was found where the first three digits of the pseudo SSN were zeros. (i.e., 000xxxxxx). Researchers should exclude these pseudo SSNs from their study until the problem is fixed. We listed the scrambled pseudo SSNs and their associated cost statistics in Table 5.21. Among the 13,203 pseudo SSNs, 13,156 were from laboratory (clinic stop=108), 29 from observation bed (clinic stop=290), 17 from employee health (clinic stop=999), and one from X-Ray (clinic stop=105). Please notice that the number of pseudo records was out of 65 million total records. Because they make up such a small fraction of the total data set, pseudo records should not have any significant impact on cost statistics such as means. However, if we calculate person-level costs, the scrambled pseudo SSN 608331723 would have a \$19 million cost in FY2002, which should be excluded.

Table 5.21 Costs of Pseudo SSNs in FY2002

SCRSSN	N	Mean (\$)	Maximum (\$)	Minimum (\$)	Sum (\$)
608331723	13,183	1,464	34,871	-4,076	19,299,197
297331383	6	463	685	2	2,780
333332333	1	98	98	98	98
409336873	7	783	1370	685	5,482
627331143	6	799	1370	685	4796

*These were identified from the entire dataset, not the random sample.

5.10 Recommendations to researchers

A unique feature of the DSS outpatient extract is that it includes many services and activities that are not recorded in the NPCD outpatient database leaving the DSS outpatient extract as the single source of such data. A general strategy when evaluating this data is to compare DSS costs for selected services that are important in study with published costs for non-VA facilities. Researchers should always examine cost outliers. Cost outliers can be examined at the person or service level. For person-level cost outliers, researchers should analyze the reasonableness of high cost services. For service-level cost outliers, researchers should identify reasonable cost ranges for major services and identify outliers for each type of services, respectively. Cost

outliers should be validated and adjusted or corrected, particularly for a study with a small sample size.

Pharmacy cost is not reported in the NPCD. To estimate outpatient pharmacy costs, researchers now have three available sources: 1) the outpatient pharmacy file in the DSS NDE, 2) the Pharmacy Benefits Management (PBM) database, and 3) the new DSS national prescription-level pharmacy extract. The outpatient pharmacy file in the DSS NDE provides total costs for each encounter at the outpatient pharmacy department, including labor, capital, and indirect costs of the pharmacy department; it does not report costs separately for each prescription. The PBM database has detailed information for each prescription. The PBM, however, does not include other costs in the pharmacy department. The PBM database is maintained at the VA Chicago Hines Medical Center by the PBM group. Access to PBM is not as easy as access to the DSS NDEs and medical SAS files at Austin. In June of 2003, the DSS team released its first pharmacy national extract that reports costs for each prescription as well as costs of the pharmacy department. The DSS pharmacy extract is a new data source and needs careful evaluation and validation.

Other major items that are reported by DSS, but not in NPCD, include prosthetics, hearing aids, eyeglasses, and other items provided by Denver Distribution Center, and Addiction Severity Index assessments. The DSS national extract is the only source to estimate costs for these services.

For services provided by VA outpatient clinics, researchers are likely to have some services reported only in NPCD and others only in DSS. The first thing researchers should do is to make sure these two groups of unmatched services are indeed unique services. Often the same services are recorded in different dates or clinic stops. For the records that cannot be matched in any case, researchers can always impute their costs for services reported in NPCD using the average costs of similar services in the DSS. For services only reported in DSS, researchers should make the decision on whether they should include or exclude those services based on their study perspective.

6. Summary

The results of reconciliation between the DSS NDEs and the NPCD encounter files in FY2002 were similar to that in FY2001. The inpatient discharge files achieved an almost perfect match (less than 0.3 percent unmatched records), a further improvement from FY2001. Problems remain in three areas: the bedsection admission and discharge dates in the treating specialty file, outpatient encounters recorded in the NPCD but not DSS, and cost outliers.

6.1 Bedsection admission and discharge dates

Using scrambled social security number (SCRSSN), station number (STA3N), bedsection admission dates, discharge dates, and bedsection name, we could only match 72% of bedsection stays in the two databases (65% in FY2001). Most of the mismatched stays, however, are due to two database-design inconsistencies. First, DSS does not count any day that is fewer than 24 hours whereas the NPCD does. This means that for bedsection stays that do not lead to final discharge (patients are transferred to another bedsection), the bedsection discharge date is likely to be one day earlier than that in the NPCD database (We did not check whether the discharge date of all such discharges in DSS were one day earlier than in NPCD). Second, for any stay that began before the start of the fiscal year, DSS resets the bedsection admission date to be the first day of the current fiscal year. In FY2002, 23% of the total bedsection stays were not matched by the five merging variables because of these two problems.

6.2 Missed outpatient records in DSS

There continue to be substantial differences between DSS and the NPCD outpatient data in FY2002. Using a 10-percent sample of VA patients, we found 1758 people in FY2001 and 2139 people in FY2002 were in the NPCD files, but not in the DSS outpatient extract. A further analysis of the FY2001 data from a single VISN showed most missed people had only one outpatient encounter during the year. Delays in transmitting data to DSS may be the cause of this discrepancy.

Researchers should consider two important factors when they use DSS outpatient cost estimates. First, DSS includes services and products that are either not recorded or not correctly recorded in the NPCD. For example, the NPCD file does not put all prosthetics records in the correct clinic stop; using DSS to identify prosthetics is more accurate than the NPCD. In both FY2001 and FY2002, DSS allocated about 20% of the outpatient cost to those services or items that were not in the NPCD. Second, about 9% (7% in FY2001) of the FY2002 NPCD records did not have a corresponding encounter in DSS with the NPCD flag. The DSS Technical Support Office provided a list of reasons for these differences. (see appendix). For NPCD records, the DSS extract is usually incomplete primarily due to later data entry. Researchers should impute costs for these missed services.

6.3 Cost outliers

Cost outliers can significantly affect study results, especially for studies with limited sample size. Unlike other problems, cost outliers can be detected and fixed easily. From the first DSS NDEs in FY1998, the number of cost outliers has continuously dropped over time. In general, cost outliers can be divided into two groups differentiated by the nature of the problem. The first group of cost outliers may be due to changes in costing structure (e.g., relative value, department volume, indirect cost allocation). Outliers in this type are usually fixed in the following year and

occur at different sites each year. Because it is not practical for the DSS team to detect and fix the outliers before the NDEs are created, we expect this type of outlier will occur every year. The second type of outlier is due to data entry or other problems within a local DSS team. These cost outliers usually concentrate in a few sites and may last for years. In any case, researchers should examine and correct outliers for their studies.

6.4 Value of DSS cost data to VA research

Readers should realize that the DSS national cost data extracts have a great value to VA health-care studies. They provide a national cost database containing every inpatient and outpatient encounter for all VA patients (4.5 million people in FY2002) in a relatively contemporary time period (3 to 4 months after the end of each fiscal year). Furthermore, the DSS is capable of providing more accurate cost data because it allocates personnel costs based on activity in minutes and encounter costs based on use of products. The DSS's potential contribution to VA health-care research is extremely important.

Reference

1. Yu W, Barnett P. *Reconciliation of DSS Encounter-Level National Data Extracts with the VA National Patient Care Database FY2001*. Menlo Park, CA: VA Health Economics Resource Center; October 2002. Technical report No. 4.
2. Yu W, Barnett PG. *Research Guide to Decision Support System National Cost Extracts*. Menlo Park: VA HSR&D Health Economics Resource Center; March 2002.

Appendix. Comparison of Outpatient Encounter Workload Between DSS & NPCD (Or Other Databases)

DSS Bedford Technical Support Center

Twenty conditions exist where NPCD (OPC) or other national database records are not expected to match with DSS records:

1. PHA: Records associated with pharmacy stops (e.g., stop code 160) are not tracked in Austin, so must be excluded prior to any matching of OPC and DSS. (However, if NPCD flag = Y, the encounter is included in the match.)
2. PHONE: Records associated with telephone stops in most cases are not tracked in DSS (until FY03). In the January 2003 Seattle HSR&D CBOC comparison, records associated with telephone stops accounted for 30-50% of all OPC mismatches to DSS. The omission of telephone stops in DSS is an issue for CBOCs, since 10 of 108 CBOCs provided 20% of more of all primary care encounters via telephone care. *(NB: In FY03, DSS will collect all telephone stops in NDE.)*
3. OBS: Observation encounters are reported on DSS as outpatient encounters Stop code 290-296, when these are still reported as inpatient observation cases.
4. NOSHOWS: Not in NPCD
5. DOM/PRRTP: Outpatient records associated with inpatient stays such as psychiatry rehabilitation (SARRP, PRRTP) and DOM accounted for a small percentage of all OPC mismatches, as DSS attributes the clinic visits to the inpatient stay. VAMC outpatient records associated with inpatient stays accounted from 1-6% of all VAMC mismatches in FY00 on the Seattle HSR&D Study.
6. DUPLICATES ON SAMEDAY IN SAME PRIMARY STOP CODE: Multiple encounters of a patient that occurred on the same day with the same stop code are recorded individually in Austin but are recorded as one encounter in DSS, so result in one or more “duplicates” among the Austin OPC mismatches. Elimination of these Austin “duplicates” can further reduce the number of OPC mismatches.
7. RAD: Radiology stops are assigned several different numbers in Austin OPC but are assigned encounter number for stop code 105 in DSS, which resulted in a handful of radiology mismatches in OPC that were accounted for by recoding to stop code 105. Some of these recoded Austin OPC records then matched DSS radiology stops that had been recoded to stop code 105. In FY02, this issue was resolved.
8. NON-VA ECS: A small number of OPC mismatches were related to non-VA care, such as stop codes 650-658, 681-685 for contracted nursing homes, etc. These mismatches typically accounted for less than 10 mismatches per year.
9. POSTING & DSS COSTING: A possible cause of mismatches is related to the different dates on which DSS and Austin close their books to any further revision. DSS stops making changes and collecting data on October 19 of each year, but Austin OPC continues to accept changes and new data after that date. This variation in data collection and correction could generate mismatches in DSS-OPC searches.
10. LATE ENTRY INTO NPCD: Any encounter coming to DSS after the previous month's DSS posting, will not receive costs. (Last-minute data entry each fiscal year for periods before September will not be costed.) *(NB: In FY03, DSS will collect non-costed encounters on a special NDE supplement. Therefore, these no-cost late encounters can be used to further match NPCD.)*

11. MEANS SCREEN: Some records from DSS feeds (CLI and UTIL) are present on the DSS National Outpatient File but are not captured in NPCD due to means test failures on the VistA-AAC editor. To ensure maximum matching with OPC records, add these records and merge remaining mismatches with these records by patient identifier, STA5A and clinical stop code. See the HSR&D reduced the total number of remaining OPC mismatches down from 6.7% to 5.6% of all OPC records in FY00 and from 3.8% to 3.5% in FY01.
12. MEANS: DSS collects 3.5 – 6.7% more OPC records than NPCD, as the AAC-edit for current means tests removes these from NPCD but DSS collects via CLI Flag = Y. (See above #9 for the records, which never get to NPCD.)
13. REF LAB: Reference Lab patients are reported on DSS as SSN 000123456 (Stop 108). Some may or may not be reported on NPCD by actual SSN.
14. ECS: Some VAMCs did not setup their ECS with the correct stop code, but rather report stop as “ECS.” These patient encounters go to NPCD with the correct stop code, but on DSS extracts are found under Stop Code = ECS.
15. UNK: Also some stop codes are UNKNOWN on DSS (usually from non-OR Surg procedures incorrectly setup to feed DSS the correct stop code). These encounters have a 400 Series stop code on NPCD, but are found under UNK stop on DSS.
16. TELEHOME BUDDY: From the second half of FY03, Telehome Monitoring (by non-video monitoring such as Telehome Buddies using web enablement), Telehome Monitoring-only using Stop code 584 (as non-count MAS, but count to DSS) will lead to more SSN-encounters from DSS than in NPCD)
17. 416: Effective FY04, the RN-only care of patients on the same day of surgery, traditionally entered as Stop Code 416, will be entered as non-count to MAS (and NPCD) but count to DSS. Therefore, DSS and NPCD will match on 429 Stop on day of surgery, but DSS will have more data for the pre-op preparation in Stop Code 416.
18. PRO: DSS receives records from the Prosthetic VistA Package on the day the PRO item is received by the patient. Effective some time in FY02, PRO fed all PRO orders to NPCD automatically on the date of order, by batch download at night to NPCD Stop Code 423. These 423 encounters are seen on DSS as NPCD flag – Yes.

The actual costs data is found on DSS from the PRO extract on the Date Received by the Patient, where PRO flag = Yes.
19. DDC: The Denver Distribution Center sends reports to DSS and these are entered into DSS as HCPC costs to a clinic with Stop Code DDC. All these supply-distribution per SSN encounters are filed on DSS as occurring at Stop Code DDC.
20. PTSD & HOMELESS: Effective in FY02, DSS added records from PTSD national Mental Health file. In FY03, DSS added records from the national Mental Health file (a) for Homeless patients identified on a homeless survey; and (b) for Homeless patients who have been discharged from Homeless Rehabilitation programs. None of these cost-free encounters are found on NPCD, but all are found in DSS NDE in FY03 followup.