

**Comparison of DSS Encounter-Level National Data Extracts and the VA
National Patient Care Database:
FY2004**

**Samuel King, MS, MDiv
Ciaran S. Phibbs, PhD
Wei Yu, PhD
Paul G. Barnett, PhD**

November 26, 2007

HERC Technical Report No. 23



**VA Palo Alto Health Care System
795 Willow Road (152 MPD)
Menlo Park, CA 94025
650 617-2630 *phone*
650 617-2639 *fax***

Table of Contents

Tables	iii
Executive Summary	v
1. Introduction.....	1
2. Inpatient Discharges.....	2
2.1 Data.....	2
2.2 Methods.....	3
2.3 Results of comparison.....	5
2.4 Summary of the discharge and PTF main files comparison	6
2.5 Recommendations to researchers.....	6
3. Inpatient Bedsection Stays.....	7
3.1 Data.....	7
3.2 Methods.....	7
3.3 Results of the treating-specialty comparison	9
3.4 Recommendations to researchers.....	11
4. Comparison between DSS treating specialty and DSS discharge file	12
4.1 Differences in cost between files	12
4.2 Stays only in treating specialty file.....	13
4.3 Stays with negative cost.....	14
5. Outpatient encounters	15
5.1 Number of records and costs.....	15
5.2 Multiple categories.....	19
5.3 Cost outliers	20
5.4 Outpatient encounters	24
5.5 Missed patients.....	25
5.6 Multiple clinic stop encounters in a single visit.....	27
5.7 Results of outpatient comparison.....	27
5.8 Categories of DSS records not found in NPCD.....	28
5.9 Other outpatient services in DSS but not in NPCD	32
5.10 Pseudo social security numbers	33
5.11 Recommendations to researchers.....	34
6. Summary	35
6.1 Bedsection admission and discharge dates	35
6.2 Missed outpatient records in DSS.....	35
6.3 Cost outliers	36
6.4 Value of DSS cost data to VA research.....	36
Reference	36
Appendix. Comparison of Outpatient Encounter Workload Between DSS & NPCD (Or Other Databases).....	37

Tables

Table 2.1 Number of discharges in DSS NDE and PTF inpatient discharge files FY2003 – FY2004	2
Table 2.2 Number of records in PTF inpatient discharge files after deletions by fiscal year FY2003-FY2004	4
Table 2.3 Number of discharges in DSS NDE and PTF discharge files after excluding community nursing home stays and duplicates FY2003 – FY2004	4
Table 2.4 Comparison of DSS NDE discharge file with PTF main files FY2003	5
Table 2.5 Comparison of DSS NDE discharge file with PTF main files FY2004	5
Table 3.1 Records in DSS treating specialty and PTF bedsection files FY2003-FY2004	7
Table 3.2 DSS treating specialty file adjustments for the comparison FY2003 - FY2004	8
Table 3.3 Number of records in PTF bedsection files after deleting duplicates: FY2003 – FY2004	9
Table 3.4 Equivalent file names in PTF and DSS treating specialty files	9
Table 3.5 Comparison of treating specialty stays between DSS and PTF files FY2003	10
Table 3.6 Comparison of treating specialty stays between DSS and PTF files FY2004	11
Table 3.7 Differences in bedsection-in and out dates between DSS and PTF files among the unmatched records FY2004	11
Table 4.1 Comparison of DSS national extract discharge and DSS treating specialty file FY2003 - FY2004	12
Table 4.2 Number of stays in DSS treating specialty file not in DSS discharge file FY2003 - FY2004	13
Table 5.1 DSS national outpatient extract FY2003-FY2004	15
Table 5.2 DSS outpatient clinic extract records classification	16
Table 5.3A Number of DSS outpatient records and total costs for each data category FY2003 and FY2004 (normal cost data only)	17
Table 5.3B Number of DSS outpatient records and total costs for each data category FY2004 (low cost data only and normal cost and low cost data)	18
Table 5.4A DSS encounters and costs with two categories FY2004 (includes normal cost data only)	19
Table 5.4B DSS encounters and costs with two categories FY2004 (includes low cost data only)	19
Table 5.4C DSS encounters and costs with two categories FY2004 (includes normal cost and low cost data)	20
Table 5.5 Cost outliers in the DSS outpatient extract FY2003-FY2004	20
Table 5.6 Stations with three or more cost outliers in the DSS outpatient clinic extract FY2003	21
Table 5.7 Stations with three or more cost outliers in the DSS outpatient clinic extract FY2004	21
Table 5.8 Stations with three or more cost outliers in the DSS outpatient pharmacy extract FY2003	22
Table 5.9 Stations with three or more cost outliers in the DSS outpatient pharmacy extract FY2004	22
Table 5.10 The top 10 clinic stops with outliers in the DSS outpatient clinic extract FY2003....	23
Table 5.11 The top 10 clinic stops with outliers in the DSS outpatient clinic extract FY2004....	23
Table 5.12 DSS records with the top 10 clinic cost outliers FY2004 (includes normal cost data only)	24
Table 5.13 DSS records with the top 10 pharmacy cost outliers FY2004 (includes normal cost data only)	24

Table 5.14 Patients and outpatient encounters in the DSS outpatient extract and NPCD outpatient SE file FY2003 and FY2004.....	25
Table 5.15 Number of patients who were recorded in only one database	26
Table 5.16 Comparison of outpatient records between the SE and the DSS NPCD category FY2003 and FY2004.....	27
Table 5.17 Other reasons for DSS non-matched outpatient records FY2004.....	29
Table 5.18A Clinic stops with more than 100,000 records (stratified by category) among DSS-only non-matched outpatient records FY2004 (includes normal cost data only)	30
Table 5.18B Clinic stops with more than 100,000 records (stratified by category) among DSS-only non-matched outpatient records FY2004 (includes low cost data only)	31
Table 5.18C Clinic stops with more than 100,000 records (stratified by category) among DSS-only non-matched outpatient records FY2004 (includes normal cost and low cost data)	32
Table 5.19 Costs of Pseudo SSNs in DSS in FY2004	33

Executive Summary

The DSS national cost data extracts are of great value to researchers and VA health-care studies. They provide a national cost database containing every inpatient and outpatient encounter for all VA patients (4.9 million patients in FY2004) in a relatively contemporary time period (3 to 4 months after the end of each fiscal year). Furthermore, the DSS is capable of providing accurate cost data because it allocates personnel costs based on activity in minutes and encounter costs based on use of products. Prior to 1998, VA researchers did not have any source of data on the cost of VA health care encounters. Because of the need to assess patient care costs, VA implemented DSS, a state of the art activity-based cost allocation system that had been widely adopted in the private sector. At first, each hospital kept its own DSS database, and it was difficult for researchers to access these data. Starting in 1999, national files of DSS were created, providing researchers with a comprehensive source of data on the cost of all VA provided health care. DSS uses local supplied data on cost, workload, and relative value.

The Health Economics Resource Center (HERC) also creates national files with an estimate of the cost of VA health encounters. The HERC estimates are based on the assumption that the costs of VA services are equal to the hypothetical Medicare reimbursements, adjusted for VA cost experience. Since DSS does not make this strong assumption, DSS data can be used to judge the relative efficiency of VA providers in producing different health care products. DSS methods allow its cost estimates to reflect variations in resource use that cannot be captured by the procedure and diagnosis codes that HERC relies on to make its cost estimates.

For many research purposes, the investigator also needs data such as diagnoses (ICD-9 codes) that are not in the DSS cost extracts, but are contained in the VA inpatient (PTF) and outpatient (NPCD) care files. The purpose of this HERC Technical Report is to provide researchers with information about how the DSS data compare with the PTF and NPCD files so that they can make the best use of the DSS cost estimates in their research projects. Towards this end, this report summarizes the results of linking the DSS NDEs with the National Patient Care Database (NPCD) outpatient file and Patient Treatment Files (PTF) in FY2004. Results of the FY2003 comparison are included for reference. As the quality of the DSS, PTF and NPCD data constantly improves, we expect that discrepancies between the DSS data and the PTF and NPCD files will also decrease over time.

Inpatient discharges The number of inpatient discharges recorded in the NDE discharge file and the three PTF discharge files were nearly identical. Although the number of unmatched DSS NDE discharges increased from 62 discharges in FY2003 to 135 discharges in FY2004, the percentage of unmatched records remained well below 1%.

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). We compared the DSS NDE treating specialty file and the PTF bedsection file. We identified records that had the same scrambled social security number (SCRSSN), station number (STA3N), inpatient bedsection admission date, discharge date, and bedsection number in the two files. We found that 73.3% DSS bedsection stays matched exactly by those five variables with the PTF database in FY2004, an increase of 0.1% from FY2003. These files have different rules for setting bedsection admission and discharge dates, however. Admission or discharge dates sometimes

differed by one day; patients admitted in a prior fiscal year were sometimes assigned the first day of the fiscal year as their admission date. When we adjusted for these differences, the percentage of matches increased sharply, as the PTF files included 97.8% of the stays in the DSS treating specialty file. The DSS treating specialty file included 99.7% of the stays recorded in the acute care (PB) PTF bedsection file and 97.4% of the stays in the extended care (XB) PTF bedsection file.

DSS treating specialty vs. DSS discharge When we compared the DSS treating specialty file with the DSS discharge file, we observed some stays that were contained in the discharge file but not in the treating specialty file. Between FY2003 and FY2004, differences of this type continued to decrease sharply (from 221 to 16). Another type of difference was stays contained in both the discharge and treating specialty files but with costs that differed by more than each of \$100, \$1000 and \$5000. Between FY2003 and FY2004, differences of this type decreased sharply (e.g., from 2,367 in FY2003 to 539 in FY2004 for stays in both files and for costs that differed by more than \$100).

Outpatient encounters The DSS outpatient extract is designed to include many outpatient services that are not recorded in the NPCD database. For example, DSS identified 21% more outpatient services (such as prosthetics and addiction severity index tests) that were not recorded in the Austin NPCD. 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. Our comparison results showed that almost all records with an NPCD flag in the DSS outpatient file had corresponding records in the NPCD SE file. For example, in FY2003, 9.4% of the records in the NPCD SE file did not have corresponding DSS records with the NPCD flag equal to "Y". In FY2004, we expanded our analysis by using a file that contains outpatient encounters that were assigned low cost by DSS ("low cost" data refer to those outpatient encounters that are either not assigned costs or assigned costs between -\$1 and \$1). In FY2004, the percentage of non-matched records decreased to 8.8% for normal cost data. However, when DSS records consisted of normal cost and low cost data, the percentage of non-matched NPCD SE records decreased sharply to 0.6%. The significance of this finding is that by including low cost DSS encounters in the FY2004 comparison between the DSS and NPCD databases, outpatient utilization thought to be missing from DSS was found, and almost all outpatient care was found to be reported in DSS. Compared with FY2003, the number (and percentage) of DSS records simultaneously classified into any two categories of outpatient activity increased sharply (from 483,522 records in FY2003 to 5,245,632 records in FY2004), and aggregated normal costs of the records increased from \$95,564,540 in FY2003 to \$398,149,399 in FY2004. In addition, the number of NOSHOW records suddenly dropped from 4,467,576 (in FY2003) to 0 (in FY2004). The large number of NOSHOW records seems to have "shifted" from the NOSHOW category (4,467,576 records in FY2003) to the CLI+NOSHOW category (4,524,055 records in FY2004). Excluding CLI+NOSHOW encounters in DSS would further make the DSS like the NPCD database.

Outpatient cost outliers We identified any clinic encounter that cost \$100,000 or more as an outlier. In FY2004, there were 47 outliers (86 in FY2003) in the pharmacy file and 121 outliers (123 in FY2003) in the clinic file. Most (37%) of the non-pharmacy outliers involved prosthetics services.

Conclusion The DSS national data extracts can be linked almost perfectly with the VHA discharge and outpatient data sets after adjustments in database design are made. For example, with these adjustments, inpatient stays in the FY2004 DSS NDEs matched almost perfectly with corresponding records in the PTF. In addition, in a comparison of the DSS treating specialty file with the DSS discharge file, discrepancies decreased sharply for stays contained in both the discharge and treating specialty files but with costs that differed by more than each of \$100, \$1000 and \$5000. For outpatient services, the two databases differed largely in design. More than 90% of the records in the NPCD event file were linked to the DSS for cost information whereas DSS allocated 21% of outpatient cost to services other than those recorded in NPCD. The correspondence between the DSS and NPCD databases (particularly for records but also for patients) sharply improved when DSS records consisted of normal cost and low cost data.

1. Introduction

This report presents the results of a comparison 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 FY 2003 and FY 2004.

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 the use of these data by central office staff, planners, and VA researchers, comprehensive encounter-level extracts of DSS production data have been created and placed at the Corporate Franchise Datacenter (CFD) 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 CFD, include additional information about these encounters such as patient demographic characteristics and the diagnoses and procedures associated with each encounter. These demographic and clinical data are not included in the DSS national data extracts.

HERC first compared 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 extracts.³ HERC identified some discrepancies between the two databases, especially for outpatient care, in the FY2000 comparison. HERC also compared the FY2001¹, and FY2002² and FY2003 data with detailed analysis of outpatient data. The following chapters report the comparison results for inpatient discharge, inpatient treating specialty, and outpatient files for FY2003 and FY2004.

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 databases and reports comparison of FY2004 discharges and the matching results.

2.1 Data

The DSS and PTF discharge files can be joined by the key variables, SCRSSN, ADMITDAY, STA3N and DISDAY. 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 grouped into three files: 1) the PM file for discharges from hospital main bedsections, 2) the XM file for discharges from VA nursing homes, domiciliary and other residential health care bedsections, and 3) the PMO file for discharges from VA hospital observational bedsections.

Table 2.1 lists the number of discharges reported in the DSS and PTF files in FY2003 and FY2004. Differences in the total numbers of records between DSS and PTF increased slightly from 55,614 in FY2003 to 57,240 in FY2004. In addition, between FY2002 and FY2003, the change in the number of discharges reported in both DSS and PTF was notably less than between FY2003 and FY2004. For example, between FY2002 and FY2003, the change in the number of discharges (1,466) reported for the DSS NDE discharge file was notably less than the change in the number of discharges (10,982) reported between FY2003 and FY2004. A similar trend occurred in the PTF files: For example, between FY2002 and FY2003, the change in the number of discharges (2,145) was notably less than the change in the number of discharges (12,608) reported between FY2003 and FY2004. As DSS began classifying (in FY2002) all observational beds as outpatient encounters, discharge records from the PMO file are excluded from the comparison process in the methods section below.

Table 2.1 Number of discharges in DSS NDE and PTF inpatient discharge files FY2003 – FY2004

Data source	FY2003	FY2004
DSS NDE discharge file	632,421	643,403
PTF discharge files	688,035	700,643
PTF main acute discharge file (PM)	560,881	571,389
PTF main non-acute discharge file (XM)	80,163	80,809
PTF main observation discharge file (PMO)	46,991	48,445

2.2 Methods

Merging variables

The DSS and PTF 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).

Community nursing home stays deletions

Before linking records from the two databases, we first excluded discharges from community nursing homes that were recorded in the PTF XM file because the DSS inpatient files do not include discharges from community nursing homes.¹ The community nursing home stays were identified by the variable STATYP=42 in the XM file. There were 7,849 discharges from community nursing homes in FY2004.

Deletion of duplicates

Records with the same values for the SCRSSN, STA3N, ADMITDAY and DISDAY variables were considered to be duplicates. There were no duplicate records in DSS.

After removing the community nursing home stays from the XM file, we checked for duplicate records within and between the PM and XM files. In FY2004, there were 228 duplicates in the PM file and one duplicate in the XM file. In addition, 45 records in the XM file were also found in the PM file. These XM file records were considered to be duplicates and were deleted (from the XM file).

Table 2.2 lists the number of discharges in the PTF files and the final number of records for comparison with the DSS database after deleting community nursing home stays and duplicate records within either a single file or across two or more PTF files. Between FY2003 and FY2004, the number of deleted duplicate records increased 311% (from 88 to 274). In contrast, between FY2002 and FY2003, the number of deleted duplicate records decreased 37% (from 140 to 88). In spite of the sharp increase in deleted duplicate records between FY2003 – FY2004 (compared with the interval FY2002 - FY2003), the number (274) of deleted duplicate records in FY2004 represents a small percentage (0.04%) of the total number (644,075) of records for comparison.

¹ However, DSS records community nursing home stay data in the outpatient cost (OPAT) file for each fiscal year.

**Table 2.2 Number of records in PTF inpatient discharge files after deletions by fiscal year
FY2003-FY2004**

	Original	Community nursing homes deletions	Duplicate deletions	Records for comparison
FY03 Total	641,044	7,840	88	633,116
PTF main acute discharge file (PM)	560,881	-	44	560,837
PTF main non-acute discharge file (XM)	80,163	7,840	44	72,279
FY04 Total	652,198	7,849	274	644,075
PTF main acute discharge file (PM)	571,389	-	228	571,161
PTF main non-acute discharge file (XM)	80,809	7,849	46	72,914

Note: Duplicates deletion includes those records within the same data set or across two or more data sets.

Table 2.3 lists the number of records in the DSS and PTF databases after excluding community nursing home stays and duplicate records. As DSS began classifying (in FY2002) all observational beds as outpatient encounters, we did not include PMO records in our FY2003 and FY2004 summaries.

**Table 2.3 Number of discharges in DSS NDE and PTF discharge files after excluding
community nursing home stays and duplicates FY2003 – FY2004**

Data source	FY2003	FY2004
DSS NDE discharge file	632,421	643,403
PTF discharge files	633,116	644,075
PTF main acute discharge file (PM)	560,837	571,161
PTF main non-acute discharge file (XM)	72,279	72,914
PTF main observation discharge file (PMO)	-	-
Difference between number of discharges in the DSS NDE discharge file and the PTF discharge file	695	672

2.3 Results of comparison

Tables 2.4 and 2.5 summarize the comparison of discharge files for FY2003 and FY2004, respectively.

As 15 observational bedsection discharges were unintentionally included in the FY2003 DSS database, we included the PMO file in the comparison of FY2004 discharges. One hundred twenty-nine records from the PMO file matched with records in the DSS discharge file in FY2004 (Table 2.5).

Table 2.4 Comparison of DSS NDE discharge file with PTF main files FY2003²

	DSS NDE (n=632,421)	PTF Main (n=633,116 ³)		
		PM (n=560,837)	XM (n=72,279)	PMO (n=46,913)
Merge with all four variables	632,359 (99.99%)	560,361 (99.92%)	71,998 (99.61%)	15 (0.03%)
Unmatched records	62	476	281	46,898 ⁴

Table 2.5 Comparison of DSS NDE discharge file with PTF main files FY2004⁵

	DSS NDE (n=643,403)	PTF Main (n=644,075)		
		PM (n=571,161)	XM (n=72,914)	PMO (n=48,313)
Merge with all four variables	643,268 (99.98%)	570,446 (99.87%)	72,719 (99.73%)	129 (0.27%)
Unmatched records	135	715	195	48,184

² In Table 2.4, the comparison of the DSS NDE discharge file with the PTF main files is two-way; that is, in the DSS NDE column, the 632,359 records are those records that merged on all four variables (SCRSSN, STA3N, ADMITDAY and DISDAY) with at least one of the PM, XM or PMO data sets. The 62 unmatched records in the DSS NDE column are those records that did not merge on all four variables (SCRSSN, STA3N, ADMITDAY and DISDAY) with at least one of the PM, XM or PMO data sets. Similarly, for each of the PM, XM and PMO columns, the 560,361, 71,998 and 15 (respective) records are those records that merged on all four variables (SCRSSN, STA3N, ADMITDAY and DISDAY) with the DSS NDE discharge file. In addition, for each of the PM, XM and PMO columns, the 476, 281 and 46,898 (respective) unmatched records are those records that did not merge on all four variables (SCRSSN, STA3N, ADMITDAY and DISDAY) with the DSS NDE discharge file.

³ PMO records are excluded.

⁴ The number (46,928) of FY2003 unmatched PMO records is 30 more than the number of records (46,898) reported above. This difference may result from a possible updating of the PMO database with new data.

⁵ The comparison in Table 2.5 was managed in a “two-way” manner similar to that described in Table 2.4.

In general, the two data sets compared quite well in FY2003-FY2004, and the number of unmatched records decreased for the XM file. While the number of unmatched records increased for the DSS and PM files, the percentage of unmatched records never rose above 1% for these (and the XM) files. 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 a 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.

2.4 Summary of the discharge and PTF main files comparison

The number of unmatched DSS discharges increased from 62 in FY2003 to 135 in FY2004. The PTF main files had more unmatched discharges than the DSS NDE file, and the XM unmatched discharges decreased from 281 in FY2003 to 195 in FY2004. However, the PM unmatched discharges increased from 476 in FY2003 to 715 in FY2004. Most of the unmatched discharges in the DSS database are 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).

2.5 Recommendations to researchers

Researchers should be able to link the two data sets accurately for inpatient discharges. Although PTF contains a few more discharges than the DSS NDE file, it is unlikely that the missed discharges will affect the results of most studies. Researchers could use the HERC average cost estimates to estimate the cost of the discharges that are found only in the 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 and then combined with the PTF bedsection files to create a file with cost and clinical information for each bedsection stay.

3.1 Data

Bedsection stays in the DSS NDE treating specialty extract were compared with those in the PTF bedsection files. 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 FY2003 and FY2004, DSS treated stays in observation bedsections as outpatient care. Therefore, records in observation bedsections were excluded from our FY2003 and FY2004 comparisons. Table 3.1 displays the number of records for each file used for the comparison in the two data sets.

Table 3.1 Records in DSS treating specialty and PTF bedsection files FY2003-FY2004

Source	FY2003	FY2004
DSS NDE treating specialty file*	1,205,406	1,218,682
PTF bedsection files, total	867,158	879,218
PTF acute bedsection file (PB)	736,136	745,601
PTF non-acute bedsection file (XB)	84,030	85,167
PTF observation bedsection file (PBO)	46,992	48,450

* The DSS NDE treating specialty file reports bedsection stays 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 either (1) two or more stays occur in different months or (2) the start and end dates of the stay span 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 stays that are not discharged or transferred by the end of the 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 data sets, 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 and methods used for the comparison. We use the term bedsection admission date for the date a patient is admitted or transferred to a bedsection. The term bedsection discharge date refers to the date a patient is discharged or transferred to another bedsection.

Census records

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 compare 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” (indicating that a patient was still in the hospital at the end of the fiscal year) 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 monthly cost information for each bedsection stay. If either (1) two or more stays occur in different months or (2) the start and end dates of the stay span across two or more months, 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.” The main and bedsection files were merged by STATYP to identify stays in community nursing homes.

Duplicated records

Records within a file with the same values in the five aforementioned variables (SCRSSN, STA3N, TRTIN, TRTOUT, and TRTSP) were considered to be duplicates and were removed before the comparison.

Tables 3.2 and 3.3 summarize (for FY2003 – FY2004) the adjustments described above for the DSS TRT and PTF bedsection files, respectively.

Table 3.2 DSS treating specialty file adjustments for the comparison FY2003 - FY2004

Adjustments	FY2003 No. of Records	FY2004 No. of Records
Original	1,205,406	1,218,682
Non-discharged stays (census records)	150,523	152,153
Multiple records per bedsection stay	235,735	233,508
Records for comparison (unique bedsection stays)	819,148	833,021

Table 3.3 Number of records in PTF bedsection files after deleting duplicates: FY2003 – FY2004

Adjustments	Original	Community nursing homes	Duplicates	Records for comparison
FY2003 Total	820,166	7,841	77	812,248
Acute bedsection (PB)	736,136	-	73	736,063
Extended bedsection (XB)	84,030	7,841	4	76,185
FY2004 Total	830,768	7,850	311	822,607
Acute bedsection (PB)	745,601	-	303	745,298
Extended bedsection (XB)	85,167	7,850	8	77,309

Note: observation stays were not included because they were not included in DSS in FY2003-FY2004.

Variables used in the merge statement

The DSS NDE and PTF bedsection files were merged on 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.

Note 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 data sets.

Table 3.4 Equivalent file names in PTF and DSS treating specialty files

PTF Names	Treating Specialty Names
BSINDAY	TRTIN
BSOUTDAY	TRTOUT
BEDSECN	TRTSP

3.3 Results of the treating-specialty comparison

Tables 3.5 and 3.6 summarize the treating-specialty comparison between the DSS NDE and PTF bedsection files for FY2003 and FY2004, respectively. The summaries reveal that the frequency and percentage of unmatched records in both FY2003 and FY2004 for both DSS NDE and PTF data was similar. While the number of unmatched DSS records increased slightly (from 219,357 in FY2003 to 222,643 in FY2004), the percentage of unmatched DSS records decreased slightly (from 26.8% in FY2003 to 26.7% in FY2004). Both the number and percentage of unmatched

PB records decreased slightly (from 195,438 or 26.6% in FY2003 to 194,496 or 26.1% in FY2004). Table 3.7 summarizes the differences in bedsection admission and discharge dates between DSS and PTF files among the unmatched records.

For both fiscal years, the DSS and PTF data differed as follows: In contrast with PTF data, the DSS treating specialty file generally sets the bedsection admission dates to the first day of the fiscal year (e.g., October 1, 2002 for FY2003) for patients who were admitted before the first day of the current fiscal year. For both FY2003 and FY2004, this resetting (or characterization of) bedsection admission dates occurred with an 85% frequency.⁶ When we adjusted for this difference, we obtained (of the total unmatched FY2004 stays) matches ranging from 5% (for PTF PB data) to 10% (for DSS data) to as high as 71% (for PTB XB data). The high percentage (71%) of matches for XB data is reasonable because stays in the XB file (which contains stays in nursing homes, domiciliary facilities, and residential programs) are much longer than in the PB file.

After adjusting for the change in bedsection admission dates in DSS data, we combined the PTF and DSS data by (1) requiring a match in the scrambled social security number, the station number and the bedsection number and (2) by allowing a one-day difference in admission or discharge dates to be considered a successful match. The comparison attained after making this adjustment was very high. As Table 3.5 reveals (for FY2003 data), the percentage of unmatched records for PB, DSS and XB data was 0.7%, 1.9% and 2.2%, respectively. Table 3.6 reveals that for FY2004, the comparison was similar, as the percentage of unmatched records for PB, DSS and XB data was 0.3%, 2.2% and 2.6%, respectively. With further examination and adjustments (e.g., perhaps on scrambled social security numbers), most of the unmatched records could likely be linked.

Table 3.5 Comparison of treating specialty stays between DSS and PTF files FY2003⁷

	DSS NDE (n=819,148)	PTF Bedsection Files (n=812,248)	
		PB (n=736,063)	XB (n=76,185)
Merge with all five variables	599,810	540,625	59,185
Unmatched records (percent of the total)	219,357 (26.8%)	195,438 (26.6%)	17,000 (22.3%)
Unmatched number of records after adjustments for inconsistency in bedsection admission and discharge dates (percent of the total)	11,294 (1.9%)	3,577 (0.7%)	1,300 (2.2%)

Note: PBO records were not compared in FY2003 because DSS classified observation stays as outpatient care.

⁶ Another difference between the two files is that 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. We do not (for FY2004) include a separate row for this adjustment (in Table 3.7) because we found that the PTF and DSS data could be compared to a very high degree by making the two adjustments discussed above in section 3.

⁷ The comparisons in Tables 3.5 and 3.6 were managed in a “two-way” manner similar to that described in Table 2.4.

Table 3.6 Comparison of treating specialty stays between DSS and PTF files FY2004

	DSS NDE (n=833,021)	PTF Bedsection Files (n=822,607)	
		PB (n=745,298)	XB (n=77,309)
Merge with all five variables	610,386	550,802	59,584
Unmatched records (percent of the total)	222,643 (26.7%)	194,496 (26.1%)	17,725 (22.9%)
Unmatched number of records after adjustments for inconsistency in bedsection admission and discharge dates (percent of the total)	13,453 (2.2%)	1,815 (0.3%)	1,537 (2.6%)

Note: PBO records were not compared in FY2004 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 FY2004

	DSS TRT	PTF PB	PTF XB
Unmatched number of records	222,643	194,496	17,725
Matched by setting BSINDAY as 10/1/2002 for stays admitted before 10/1/2002 (percent of the total unmatched stays)	22,373 (10.0%)	9,793 (5.0%)	12,580 (71.0%)
Unmatched number of records after adjustments for inconsistency in bedsection admission and discharge dates (percent of the total unmatched records)	13,453 (6.0%)	1,815 (0.9%)	1,537 (8.7%)

3.4 Recommendations to researchers

If researchers need to link bedsection stays in the DSS and PTF data sets, 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 adjust as explained above to conduct a match. For the 2% unmatched DSS stays, researchers may link them in the two data sets by matching on SCRSSN, bedsection number and on any combination of two of the three remaining variables. For the very few unmatched records, manual examination is necessary.

4. Comparison between DSS treating specialty and DSS discharge file

To validate whether stays and costs reported in the DSS treating specialty file are consistent with stays reported in the DSS discharge file, we compared these two files for FY2004. The DSS treating specialty file contains records for only that part of the bedsection stay that took place during the fiscal year. If a stay began in a previous fiscal year, that part of the stay is not in the current fiscal year. Rather, it is in the earlier year's DSS treating specialty file. If a bedsection stay lasts more than a single fiscal period (month), there will be multiple records for the same stay. The discharge file, however, provides one record for each hospital stay that ended during the fiscal year. Accordingly, 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 (from the treating specialty file), because such stays are not reported in the discharge file. In addition, stays that began before the first day of the fiscal year (10/1/2003 for FY2004) were excluded (from both the discharge and treating specialty files) because costs were included in the discharge file but not in the treating specialty file.

Table 4.1 shows how these sources compared. Between FY2003 and FY2004, discrepancies decreased sharply for stays in the discharge file but not in the treating specialty file. Discrepancies also decreased sharply for stays in both files, but with costs that differed by more than \$100, \$1,000 and \$5,000, respectively.⁸

Table 4.1 Comparison of DSS national extract discharge and DSS treating specialty file FY2003 - FY2004

Problem	FY2003	FY2004
Stays in the discharge file but not the treating specialty file	221	16
Stays in the treating specialty file but not in the discharge file	5,619	6,728
Stays in both files, but with costs that differed by more than \$100	2,367	539
Stays in both files, but with costs that differed by more than \$1,000	1,392	436
Stays in both files, but with costs that differed by more than \$5,000	440	263

4.1 Differences in cost between files

Referring to Table 4.1, for FY2004, 539 stays had differences in cost of at least \$100 between the treating specialty and discharge files. The cost was higher in the discharge file for 94 records; it was higher in the treating specialty file for 445 records (this is a sharp decrease from

⁸ Interestingly, between FY2002 and FY2003, discrepancies increased sharply for stays in both files, but with costs that differed by more than each of \$100, \$1,000 and \$5,000, respectively.

FY2003 when the cost was higher in the treating specialty file for 2310 records). In FY2004, no station represented more than 6% of the stays with cost differences of at least \$100 (this contrasts sharply with FY2003 when fully 1736 [or 73%] of the 2367 stays with cost differences of at least \$100 happened at station number 460 [Wilmington]). Station numbers 512 (Baltimore), 640 (Palo Alto, CA.) and 657 (St. Louis, MO.) were each responsible for 6% of the 445 stays with cost differences of at least \$100. In addition, these three stations (respectively) had 31, 34 and 33 stays with cost differences of at least \$100. These same three stations were also responsible for the highest percentage of stays with cost differences of at least \$1,000 and \$5,000. Yet no single station was responsible for more than 10% of the stays with cost differences of at least either \$1,000 or \$5,000

A few records had noteworthy differences in cost, including one record in which the cost in the treating specialty file exceeded the cost in the discharge file by \$216,765 (this represents a sharp decrease from FY2003, when the maximum difference was \$1,488,189).

4.2 Stays only in treating specialty file

For FY2004, of the 6,728 stays in the treating specialty file that are not in the discharge file, about one-third (34%) involved stays at five medical centers (see Table 4.2). This contrasts with FY2003, when nearly half (48%) involved stays at five medical centers.

Table 4.2 Number of stays in DSS treating specialty file not in DSS discharge file FY2003 - FY2004

Top 5 stations with missing stays in FY2003

STA3N	STA3N Label	Frequency
586	JACKSON	896
672	SAN JUAN	728
664	SAN DIEGO	445
520	BILOXI	317
646	PITTSBURGH-UNIV DR	317

Top 5 stations with missing stays in FY2004

STA3N	STA3N Label	Frequency
586	JACKSON	785
520	BILOXI	517
642	PHILADELPHIA	435
549	DALLAS	270
671	SAN ANTONIO	266

For FY2004, stations 549, 642 and 671 were newly added to the list of top 5 problem stations. Previously, they ranked #21, #39 and #26, respectively. Stations 646, 664 and 672, which were among the top 5 with problems in FY03, had far fewer mismatches in FY04.

4.3 Stays with negative cost

In FY2004, 28 stays were assigned a negative cost. All of these stays had negative costs in both files, and for each of the 28 stays, the negative costs (when rounded to the nearest dollar) were identical. One stay was assigned a cost of -\$15,739 (this represents a sharp change from FY2003, when the maximum negative cost was -\$752,320).

5. Outpatient encounters

The DSS extracts have been reorganized in order to decrease processing time and CPU costs of those accessing these data. The reorganization occurred in spring 2007 and impacts data beginning with FY2004 (see HERC Bulletin 7, Issue 3). Comments in this chapter about the organization of the DSS outpatient extract are true up until just before the DSS extracts were reorganized in spring 2007.

The DSS outpatient extract contains services recorded in the NPCD. The DSS outpatient file is also designed to include many outpatient services that are not recorded in the NPCD database. 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. We first identified which records in the DSS outpatient extract used NPCD as their data source. Then we matched these DSS records with records in the NPCD SE file.

5.1 Number of records and costs

For FY2003 and FY2004, the DSS outpatient extract consisted of two data sets, one for outpatient pharmacy data and one for outpatient clinic data. However, for FY2004, each of these two data sets has been separated into a normal cost and low cost database. “Low cost” data refer to those outpatient encounters that are either not assigned costs or assigned costs between -\$1 and \$1. “Normal cost” data refer to those outpatient encounters that are assigned all other dollar amount costs. For each fiscal year, each data set includes four files separated by VISNs: VISN 1 – VISN 5, VISN 6 – VISN 10, VISN 11 – VISN16, and VISN 17 – VISN 22. The outpatient pharmacy extract contains VA pharmacy utilization and the outpatient clinic extract includes all other outpatient encounters, including care not included in the NPCD. The number of records and total costs are listed in Table 5.1. Note that for FY2004, Table 5.1 lists separately the number of records and total costs for all outpatient encounters that (1) excludes low cost data and (2) includes normal cost and low cost data.

Table 5.1 DSS national outpatient extract FY2003-FY2004

File	FY2003	FY2004 (includes normal cost data only)	FY2004 (includes normal cost and low cost data)
Number of records			
Outpatient Clinic	68,468,009	74,147,244	83,311,986
Outpatient Pharmacy	57,670,780	62,761,896	63,151,949
Costs			
Outpatient Clinic	\$11,163,074,376	\$12,388,328,842	\$12,388,633,598
Outpatient Pharmacy	\$3,789,599,139	\$4,392,350,189	\$4,392,564,030

DSS allocated costs to outpatient activities recorded in the VISTA system. NPCD records only encounters with providers. DSS is more complete, recording other types of utilization. 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 Tables 5.3A and 5.3B summarize the number of records and costs for each category in the FY2003 and FY2004 outpatient clinic extract. Note that for FY2004, Tables 5.3A and 5.3B list the number of records and total costs for all outpatient encounters that (1) excludes low cost data, (2) excludes normal cost data, and (3) includes normal cost and low cost data.

Table 5.2 DSS outpatient clinic extract records classification

Flag Variable	Description
PRE	Outpatient pharmacy clinic utilization records
NOSHOW	Outpatient clinic no-show records
PROS	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	<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)
Multiple	Records assigned to two of the above categories

**Table 5.3A Number of DSS outpatient records and total costs for each data category
FY2003 and FY2004 (normal cost data only)**

Category	Records			
	FY2003		FY2004 (includes normal cost data only)	
	N	%	N	%
NPCD	55,755,342	81.4	59,932,524	80.8
CLI	1,344,228	2.0	1,121,086	1.5
PROS	2,230,553	3.3	2,365,397	3.2
DDC	565,370	0.8	594,140	0.8
NOSHOW	4,467,576	6.5	0	0.0
PRE	9,792	0.0	9,470	0.0
UTIL	1,873,561	2.7	1,976,531	2.7
All Other ⁹	1,738,065	2.5	2,902,464	3.9
MULTIPLE	483,522	0.7	5,245,632	7.1
Total	68,468,009	100	74,147,244	100
Category	Cost			
	FY2003		FY2004 (includes normal cost data only)	
	\$	%	\$	%
NPCD	\$8,819,721,077	79.0	\$9,756,330,047	78.8
CLI	\$207,881,209	1.9	\$199,118,404	1.6
PROS	\$558,275,477	5.0	\$648,631,059	5.2
DDC	\$103,158,417	0.9	\$116,052,998	0.9
NOSHOW	\$255,094,941	2.3	\$0	0.0
PRE	\$12,784,201	0.1	\$13,964,256	0.1
UTIL	\$306,698,623	2.7	\$288,386,944	2.3
All Other	\$803,895,892	7.2	\$967,695,735	7.8
MULTIPLE	\$95,564,540	0.9	\$398,149,399	3.2
Total	\$11,163,074,376	100	\$12,388,328,842	100

⁹ "All other" is a group identified by excluding records marked by the seven flags.

**Table 5.3B Number of DSS outpatient records and total costs for each data category
FY2004 (low cost data only and normal cost and low cost data)**

Category	Records			
	FY2004 (includes low cost data only)		FY2004 (includes normal cost and low cost data)	
	N	%	N	%
NPCD	5,423,118	59.2	65,355,642	78.4
CLI	386,979	4.2	1,508,065	1.8
PROS	97,881	1.1	2,463,278	3.0
DDC	74,760	0.8	668,900	0.8
NOSHOW	0	0.0	0	0.0
PRE	0	0.0	9,470	0.0
UTIL	208,627	2.3	2,185,158	2.6
All Other	2,313,146	25.2	5,215,610	6.3
MULTIPLE	660,231	7.2	5,905,863	7.1
Total	9,164,742	100	83,311,986	100
Category	Cost			
	FY2004 (includes low cost data only)		FY2004 (includes normal cost and low cost data)	
	\$	%	\$	%
NPCD	\$156,804	51.5	\$9,756,486,850	78.8
CLI	\$1,170	0.4	\$199,119,575	1.6
PROS	\$14,293	4.7	\$648,645,352	5.2
DDC	\$12,630	4.1	\$116,065,628	0.9
NOSHOW	\$0	0.0	\$0	0.0
PRE	\$0	0.0	\$13,964,256	0.1
UTIL	\$9,035	3.0	\$288,395,979	2.3
All Other	\$91,640	30.1	\$967,787,375	7.8
MULTIPLE	\$19,183	6.3	\$398,168,582	3.2
Total	\$304,756	100	\$12,388,633,598	100

According to the National DSS Extract Technical Guidebook (DSS BTSO/Development, April 3, 2000), each DSS cost record is characterized by the source of data, with the highest priority for NPCD records. The services recorded in the NPCD database accounted for 79% (for both FY2003 and FY2004) of the total DSS outpatient cost. In FY2003, 9,792 pharmacy records were included in the non-pharmacy portion of the outpatient cost file for a total cost of \$13 million. For FY2004, 9,470 pharmacy records were included in the outpatient clinic file for a total cost of \$14 million. Compared with FY2003, the number (and percentage) of records simultaneously classified into two categories increased sharply in FY2004. The characteristics of multi-category records is further investigated and reported in the following section.

Prior to FY2004, the number of NOSHOW records was fairly even (for example, the number of NOSHOW records in FY2002 and FY2003 was 4,375,320 and 4,467,576, respectively). The large number of NOSHOW records seems to have “shifted” from the NOSHOW subcategory

(4,467,576 records in FY2003) to a subcategory (CLI+NOSHOW, having 4,524,055 records in FY2004) within the MULTIPLE category. Beginning in FY2006, DSS will no longer assign costs to NOSHOW visits.

5.2 Multiple categories

For normal cost data, about 7.1% of outpatient records were assigned to two categories in FY2004. Compared with FY2003, the number of multi-category records increased from 483,522 to 5,245,632 in FY2004, and costs increased from \$95,564,540 to \$398,149,399 in FY2004 (Table 5.4A).

Throughout our report of outpatient encounters, multiple-category records were included in all summaries and were treated no differently (for reporting purposes) than records classified into a single category.

Table 5.4A DSS encounters and costs with two categories FY2004 (includes normal cost data only)

Combination	SAS Value ¹⁰	No. of records	%	Costs	%
CLI+NOSHOW	NYNNYNN	4,524,055	86.2	\$261,022,089	65.6
NPCD+PROS	YNYNNNN	642,738	12.3	\$107,806,783	27.1
CLI+PROS	NYYNINN	72,673	1.4	\$14,195,970	3.6
CLI+PRE	NYNNNYN	5,137	0.1	\$12,788,877	3.2
Others	YNNNNYN	1,029	0.0	\$2,335,680	0.6
Total	YNNNNYN	5,245,632	100	\$398,149,399	100

Table 5.4B DSS encounters and costs with two categories FY2004 (includes low cost data only)

Combination	SAS Value ¹⁰	No. of records	%	Costs	%
CLI+NOSHOW	NYNNYNN	641,269	97.1	\$13,242	69.0
NPCD+PROS	YNYNNNN	18,758	2.8	\$5,936	30.9
CLI+PROS	NYYNINN	204	0.0	\$4	0.0
Total	YNNNNYN	660,231	100	\$19,183	100

¹⁰ 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."

Table 5.4C DSS encounters and costs with two categories FY2004 (includes normal cost and low cost data)

Combination	SAS Value ¹⁰	No. of records	%	Costs	%
CLI+NOSHOW	NYNNYNN	5,165,324	87.5	\$261,035,332	65.6
NPCD+PROS	YNYNNNN	661,496	11.2	\$107,812,719	27.1
CLI+PROS	NYYNINN	72,877	1.2	\$14,195,974	3.6
CLI+PRE	NYNNNYN	5,137	0.1	\$12,788,877	3.2
Others	YNNNNYN	1,029	0.0	\$2,335,680	0.6
Total	YNNNNYN	5,905,863	100	\$398,168,582	100

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 FY2003, the FY2004 data had sharply fewer outliers in the pharmacy file and slightly fewer outliers in the clinic file. The largest outlier in a single pharmacy utilization was assigned a cost of \$320,000.

Table 5.5 Cost outliers in the DSS outpatient extract FY2003-FY2004
(Total cost >= \$100,000/record)

	FY2003	FY2004
Pharmacy		
Number of records	86	47
Total cost	\$23,623,266	\$7,636,230
Maximum cost /record	\$812,737	\$324,187
Other Clinics		
Number of records	123	121
Total cost	\$27,671,254	\$28,319,744
Maximum cost / single stop visit	\$1,400,710	\$860,536

We further examined these cost outliers by medical center (STA3N) and clinic stop. In the outpatient clinic file, the number of stations with three or more outliers decreased from 13 in FY2003 to 7 in FY2004 (Tables 5.6 and 5.7). For two stations (Washington, D.C. and Houston), cost outliers occurred in both years. In the pharmacy file, the number of stations with three or more outliers decreased from 9 in FY2003 to 7 in FY2004 (Tables 5.8 and 5.9). For one station (Las Vegas), cost outliers occurred in both years.

**Table 5.6 Stations with three or more cost outliers in the DSS outpatient clinic extract
FY2003**
(Total cost >=\$100,000/record)

STA3N	STA3N Label	Frequency	Percent
612	MARTINEZ	15	12.20
580	HOUSTON	14	11.38
673	TAMPA	13	10.57
693	WILKES BARRE	11	8.94
463	ANCHORAGE	8	6.50
663	SEATTLE	5	4.07
757	COLUMBUS-IOC	5	4.07
546	MIAMI	4	3.25
688	WASHINGTON D.C.	4	3.25
534	CHARLESTON	3	2.44
589	KANSAS CITY	3	2.44
595	LEBANON	3	2.44
660	SALT LAKE CITY	3	2.44

**Table 5.7 Stations with three or more cost outliers in the DSS outpatient clinic extract
FY2004**
(Total cost >=\$100,000/record)

STA3N	STA3N Label	Frequency	Percent
523	BOSTON	51	42.15
688	WASHINGTON D.C.	14	11.57
678	TUCSON	5	4.13
580	HOUSTON	4	3.31
619	MONTGOMERY	4	3.31
659	SALISBURY	4	3.31
550	DANVILLE IL	3	2.48

**Table 5.8 Stations with three or more cost outliers in the DSS outpatient pharmacy extract
FY2003**
(Total cost >=\$100,000/record)

STA3N	STA3N Label	Frequency	Percent
553	JOHN D. DINGELL VAMC	23	26.74
637	ASHEVILLE-OTEEN	13	15.12
595	LEBANON	11	12.79
539	CINCINNATI	8	9.30
654	RENO	7	8.14
531	BOISE	5	5.81
528	BUFFALO	3	3.49
593	LAS VEGAS	3	3.49
630	NEW YORK	3	3.49

**Table 5.9 Stations with three or more cost outliers in the DSS outpatient pharmacy extract
FY2004**
(Total cost >=\$100,000/record)

STA3N	STA3N Label	Frequency	Percent
593	LAS VEGAS	10	21.28
610	NORTHERN INDIANA HCS	6	12.77
549	DALLAS	3	6.38
561	EAST ORANGE	3	6.38
580	HOUSTON	3	6.38
642	PHILADELPHIA	3	6.38
652	RICHMOND	3	6.38

Tables 5.10 and 5.11 list cost outliers by clinic stops for FY2003 and FY2004, respectively. The FY2004 data indicates that about a third each of the cost outliers in the clinic file are from prosthetics and VA-paid home/community healthcare providers. Although some items in the prosthetics category are expensive, further analysis is still needed.

Table 5.10 The top 10 clinic stops with outliers in the DSS outpatient clinic extract FY2003
(Total cost >=\$100,000/record)

CL	CL Label	Frequency	Percent
423	PROSTHETICS SVCS	65	52.85
429	OUTPAT CARE IN O.R.	16	13.01
105	X-RAY	12	9.76
681	VA-PD HOME/COMM HC	7	5.69
291	OBSERVATION SURGERY	3	2.44
323	PRIM CARE/MED	3	2.44
107	EKG	2	1.63
303	CARDIOLOGY	2	1.63
333	CARDIAC CATH	2	1.63
409	ORTHOPEDICS	2	1.63
XXX	All Others	9	7.32

Table 5.11 The top 10 clinic stops with outliers in the DSS outpatient clinic extract FY2004
(Total cost >=\$100,000/record)

CL	CL Label	Frequency	Percent
423	PROSTHETICS SVCS	42	36.84
681	VA-PD HOME/COMM HC	34	29.82
652	STATE DOM DAYS	14	12.28
429	OUTPAT CARE IN O.R.	10	8.77
651	STATE NH DAYS	6	5.26
653	STATE HOSP CARE	2	1.75
105	X-RAY	1	0.88
152	ANGIOGR CATHETERIZ	1	0.88
290	OBSERVATION MEDICINE	1	0.88
291	OBSERVATION SURGERY	1	0.88
XXX	All Others	2	1.75

Tables 5.12 and 5.13 list records that had the top 10 most expensive clinic cost and pharmacy cost outliers, respectively for FY2004 normal cost data. In the FY2004 low cost data, no station had an outlier that exceeded \$1, and no station (among the top 10 clinic or pharmacy cost outliers) appeared in either of Tables 5.12 or 5.13. In the clinic file, Station 523 (Boston) has 7 of the top 10 most expensive outliers, 2 with the same value (\$569,959). In the pharmacy file, Station 593 (Las Vegas) has 4 of the top 10 outliers, all with similar values (\$207,062 – \$207,093). In addition, the Northern Indiana HCS has 3 of the top 10 outliers, with values ranging from \$204,123 to \$241,287.

Table 5.12 DSS records with the top 10 clinic cost outliers FY2004 (includes normal cost data only)

STA3N	STA3N Label	Clinic cost outliers (\$)
539	CINCINNATI	860,536
523	BOSTON	854,939
523	BOSTON	664,953
523	BOSTON	643,351
688	WASHINGTON D.C.	640,022
501	ALBUQUERQUE	617,831
523	BOSTON	578,366
523	BOSTON	569,959
523	BOSTON	569,959
523	BOSTON	565,204

Table 5.13 DSS records with the top 10 pharmacy cost outliers FY2004 (includes normal cost data only)

STA3N	STA3N Label	Pharmacy cost outliers (\$)
658	SALEM	324,187
657	ST LOUIS-John Cochran	301,031
610	NORTHERN INDIANA HCS	241,287
463	ANCHORAGE	209,077
593	LAS VEGAS	207,093
593	LAS VEGAS	207,081
593	LAS VEGAS	207,078
593	LAS VEGAS	207,062
610	NORTHERN INDIANA HCS	204,128
610	NORTHERN INDIANA HCS	204,123

5.4 Outpatient encounters

Table 5.14, which compares the differences between the DSS outpatient clinic extract and the Austin NPCD SE file, shows that the differences between these two files are consistent in FY2003-FY2004. For both years, the number of records in the NPCD outpatient SE file is larger than the number of NPCD records in the DSS outpatient clinic extract (even when one adds the low cost DSS data to the normal cost DSS data). In addition, more patients are in the SE file than are in the DSS outpatient clinic extract. However, these differences between the NPCD outpatient SE file and the DSS outpatient clinic extract were less (particularly for records but also for patients) when the DSS outpatient clinic extract included normal cost and low cost data. For example, the number of records (72,507,936) in the FY2004 NPCD outpatient SE file was about 20% larger than the number (60,575,591) of NPCD records in the (normal cost) DSS outpatient clinic extract. This difference was only about 10% when normal cost and low cost DSS data (consisting of 66,017,467 records) was included in the DSS outpatient clinic extract.

Compared with FY2003, the number of patients in both the NPCD outpatient SE file and the (normal cost) DSS outpatient clinic extract increased by about 4%. The number of records in both the NPCD outpatient SE file and the DSS outpatient clinic extract (normal cost data only) increased by about 7% and 8%, respectively.

Table 5.14 Patients and outpatient encounters in the DSS outpatient extract and NPCD outpatient SE file FY2003 and FY2004

Data Source	FY2003	FY2004 (includes normal cost DSS data only)	FY2004 (includes low cost DSS data only)	FY2004 (includes normal cost and low cost DSS data)
Number of Patients				
NPCD outpatient SE file	4,694,401	4,876,837	4,876,837	4,876,837
DSS outpatient clinic extract (NPCD flag)	4,634,897	4,831,028	1,656,483	4,870,733
Number of Records				
NPCD outpatient SE file	68,058,725	72,507,936	72,507,936	72,507,936
DSS outpatient clinic extract (NPCD flag)	56,202,596	60,575,591	5,441,876	66,017,467

5.5 Missed patients

We also examined the number of patients who were recorded in only one database. Table 5.15 summarizes these findings for FY2003 and FY2004. For FY2004, the difference between the two databases (for the first two rows) for normal cost and low cost DSS data was far less than it was for normal cost DSS data. For example, 49,709 patients were in the NPCD SE file, but not in the (normal cost) DSS records with NPCD flag. However, this difference decreased to 10,641 patients when DSS records consisted of normal cost and low cost data. Similarly, 32,824 patients were in the NPCD SE file, but not in any (normal cost) DSS records. However, this difference decreased to 6,728 patients when DSS records consisted of normal cost and low cost data.

While the addition of low cost DSS data to normal cost DSS data caused the difference between the two databases to increase (for the last two rows of Table 5.15 in FY2004), the percentage increase of the difference was not sharp. For example, 3,900 patients were only in the (normal cost) DSS outpatient extract with the NPCD flag. This difference increased by 16% (to 4,537 patients) when DSS records consisted of normal cost and low cost data. Similarly, 242,993 patients were only in the (normal cost) DSS outpatient extract for services that are not in the NPCD. This difference increased by 4% (to 253,503 patients) when DSS records consisted of normal cost and low cost data.

For FY2003, we examined whether the 62,503 patients had any records in other categories of the DSS outpatient file. We found that 21,849 patients had some records in other DSS outpatient

categories, but 40,654 patients in the NPCD SE file had no records in any DSS outpatient categories.

We performed a similar analysis for FY2004. For normal cost DSS data, for example, we found that among the 49,709 patients who appeared in the NPCD SE file, but had no records in DSS with an NPCD flag, 16,885 had some records in DSS with other flags. However, 32,824 patients in the NPCD SE file had no records in any category of the DSS outpatient file. Researchers should understand that DSS included outpatient services that were not recorded in the NPCD SE file. For FY2004, 242,993 patients had records in DSS with only non-NPCD flags, and 57% of those patients only had records with the “No Show” flag.

Table 5.15 Number of patients who were recorded in only one database
 (Based on outpatient encounters in the DSS outpatient extract and NPCD outpatient SE file in FY2003 and FY2004)

Data Source	FY2003	FY2004 (includes normal cost DSS data only)	FY2004 (includes low cost DSS data only)	FY2004 (includes normal cost and low cost DSS data)
Patients only in the NPCD SE file				
Number of patients in the NPCD SE file, but not in the DSS records with NPCD flag	62,503	49,709	3,221,863	10,641
Number of patients in the NPCD SE file, but not in any DSS records	40,654	32,824	2,214,053	6,728
Number of patients in the NPCD SE file and in the DSS records but not in DSS records with the NPCD flag	4,631,898	4,827,128	1,654,974	4,866,196
Patients only in the DSS outpatient file				
Patients only in the DSS outpatient extract with NPCD flag	2,999	3,900	1,509	4,537
Patients only in the DSS outpatient extract for services that are not in the NPCD.	248,497	242,993	1,042,617	253,503

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 compare the two data sets, therefore, we consolidated these multiple encounters in the NPCD SE file. In FY2004, 6,144,335 (8.5%) of the 72,507,936 records in the NPCD SE file 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). In FY2004, we found 5 duplicates in the records with NPCD flags, and 3,272 duplicates in the records without NPCD flags.

5.7 Results of outpatient comparison

As previously indicated in section 5.5, the DSS outpatient extract contains many more services that are not recorded in the NPCD SE file. DSS records with NPCD flags and records in the NPCD SE file were compared. After consolidating multiple clinic encounters incurred by a patient during a single visit in both files, the two files 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 (CLNUM in the NPCD file and CL in the DSS file). Table 5.16 summarizes the comparison.

Table 5.16 Comparison of outpatient records between the SE and the DSS NPCD category FY2003 and FY2004

Data Source	FY2003	FY2004 (includes normal cost DSS data only)	FY2004 (includes low cost DSS data only)	FY2004 (includes normal cost and low cost DSS data)
Records in DSS outpatient extract with NPCD flag	56,202,596	60,575,586	5,441,876	66,017,462
Records in consolidated NPCD SE file (excluding pharmacy consultation records)	62,023,736	66,363,601	66,363,601	66,363,601
Records in both files	56,166,138	60,529,169	5,425,818	65,954,987
Records only in the DSS outpatient extract ¹¹	36,458 (0.06%)	46,417 (0.08%)	16,058 (0.30%)	62,475 (0.09%)
Records only in the NPCD SE file ¹²	5,857,598 (9.44%)	5,834,432 (8.79%)	60,937,783 (91.82%)	408,614 (0.62%)

¹¹ The denominator for the percentage calculations in this row is the number of records in the DSS outpatient extract with the NPCD flag.

¹² The denominator for the percentage calculations in this row is the number of records in the consolidated NPCD SE file (excluding pharmacy consultation records).

We found that in FY2003, 9.4% of the records in the NPCD SE file did not have a corresponding DSS NPCD record (i.e., DSS records with NPCD flag equal to “Y”). In FY2004, the percentage of non-matched records decreased to 8.8% for normal cost data. However, when DSS records consisted of normal cost and low cost data, the percentage of non-matched NPCD SE records decreased sharply to 0.6%. The percentage change of non-matched DSS records was much less sharp when comparing the normal cost data with the combined normal cost and low cost data. For example, when DSS records consisted of normal cost data, the percentage of non-matched DSS records NPCD SE records was 0.08%, which was close to the percentage (0.09%) of non-matched DSS records for the combined normal cost and low cost data.

5.8 Categories of DSS records not found in NPCD

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 among (FY2004) “DSS-only” records by first preparing (for normal cost DSS data) a file of 13,568,381 DSS records that did not have the NPCD flag. Using the four common variables (1) scrambled social security number (SCRSSN), (2) medical center identification number (3-digit numeric STA3N), (3) outpatient clinic visit date (VIZDAY), and (4) clinic stop number (CL), we then merged this file with a file of 46,417 DSS records that had the NPCD flag, but had no corresponding record in NPCD. The resulting file of “DSS-only” data consists of 13,614,689 records.

Observation bed encounters¹³

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 FY2004, 52,750 DSS-only records had an observation bed clinic stop, accounting for less than 1% of the 13,614,689 DSS-only records (Table 5.17). The 52,750 DSS-only records reported in the observation group exceed (by 9%) the number (48,450) of PTF records reported (in Table 3.1) in the observation bedsection file.

Radiology

Radiology encounters were assigned several different clinic numbers in the Austin NPCD SE file but were assigned to clinic stop 105 in DSS. According to the BTSO Memo, this issue was resolved in FY2002. However, we still observed (in FY2004) 121,725 DSS non-matched records in stop code 105, accounting for less than 1% of the DSS-only records (Table 5.17).

Contract extended care and other contract care

Contract extended care, other contract care and care attributed to the Denver Distribution Center were recorded in DSS as stop codes 640, 650-654, 656, 657, 670, 730, 731, and DDC. In FY2004, 950,511 records had these stop codes, accounting for 7.0% of the 13,614,689 DSS-only records (Table 5.17). These records include contract nursing home stays and visits by an RN to patients in contract nursing homes. In FY2002, the DSS NDE began to collect information about the length of stay for these community nursing home encounters. The data in FY2004 indicated

¹³ The statistics provided in this section (observation bed encounters) and the following two sections (radiology, and contract extended care and other contract care) are in the context of cost-only DSS data. For these three sections, the percentages of DSS-only records do not differ notably when the DSS cost data is compared with the combined cost and low cost data.

that these services cost \$815,292,668. The median cost was \$241, the maximum cost was \$854,939, and 36% of those encounters cost \$74 or less and may be associated with the RN visits.

Only 1% (or 7,855) of the 950,511 contract extended care and other contract care encounters had a negative cost. Twenty-one percent of these negative cost encounters occurred at four medical centers (Bay Pines [428 encounters], Boston [258 encounters], W. Palm Beach [341 encounters], and Spokane [591 encounters]). In FY2003, 7,596 negative cost encounters occurred.

Table 5.17 Other reasons for DSS non-matched outpatient records FY2004

Category	Frequency and percent of DSS-only records*		
	Normal cost DSS data only	Low cost DSS data only	Normal cost and low cost DSS data
Contract extended care and other contract care	950,511 (7.0%)	85,185 (2.3%)	1,035,696 (6.0%)
Radiology	121,725 (0.9%)	42,007 (1.1%)	163,711 (0.9%)
Observation bed	52,750 (0.4%)	200 (0.0%)	52,948 (0.3%)
Total DSS-only records	13,614,689 (100%)	3,738,036 (100%)	17,349,240 (100%)

* Includes records that were under the NPCD flag and only in DSS as well as DSS records under all other flags (for the “normal cost DSS data only” column, this amounts to 46,417 records).

Tables 5.3A and 5.3B display (for FY2003 and FY2004) the number of outpatient records and total costs for each data category. One of the categories in these tables was “All Other” (a category identified by records that did not have any of the seven flags). Contract extended care and other contract care visits represented a significant percentage of the total cost in the “All Other” category. For FY2003, this percentage was 91% (which represented a cost of \$729,386,660 from 900,016 records). For normal cost DSS FY2004 data, this percentage was 84% (which represented a cost of \$815,292,668 from 950,511 records).

We next explored the nature of the FY2004 “DSS-only” data in the last row of Table 5.17 (note that the normal cost and low cost DSS-only data in this row [consisting of 17,349,240 records] exceeds [by 27%] the number of records [13,614,689] for the normal cost DSS-only data). We first obtained a file of those clinic stops that occurred with a frequency of 100,000 or greater. In this file (for normal cost and low cost data), the clinic stops that occurred with the greatest frequency were prosthetics service, psychological testing, primary care/med and laboratory. For normal cost data only, these four clinic stops had a combined frequency of 6,832,799 (or fully 50% of all the clinic stops among the “DSS-only” data).

Using Tables 3.1 and 5.17, the number of PTF records reported in the observation bedsection files can be compared with the number of DSS observation records reported in the treating specialty file. When this comparison is made for FY2004, the number (48,450) of PTF records reported in the observation bedsection file is within 9% of the number (52,750) of DSS records reported in the observation group in Table 5.17.

For the file of those clinic stops that occurred with a frequency of 100,000 or greater, we obtained a tabulation of the number of clinic stops in each of the (single and multiple) categories identified in Tables 5.2, 5.3A and 5.3B. The results of this tabulation (displayed in Table 5.18A, Table 5.18B and Table 5.19C) reveal the following for normal cost data: (1) Prosthetics service (identified with the PROS [prosthetics] flag) was the highest-occurring clinic stop (with a frequency of 2,364,724 [or 17%]). For the normal cost and low cost data, psychological testing (identified with the All Other flag) was the highest-occurring clinic stop (with a frequency of 3,369,622 [or 19%]). (2) Eight clinic stop types (namely, primary care/med, mental health-ind, optometry, ophthalmology, podiatry, dermatology, urology, and substance abuse-grp) were identified in the “CLI+NOSHOW” category. These clinic stop types represented 17% of all the clinic stops among the “DSS-only” data.

Table 5.18A Clinic stops with more than 100,000 records (stratified by category) among DSS-only non-matched outpatient records FY2004 (includes normal cost data only)

CL	CL Label	Category	Frequency	Percent of DSS-only records ¹⁴
423	PROSTHETICS SVCS	PROS	2,364,724	17.4
538	PSYCHOLOGICAL TESTING	All Other ¹⁵	1,269,868	9.3
323	PRIM CARE/MED	CLI+NOSHOW	1,224,453	9.0
108	LABORATORY	UTIL	1,119,272	8.2
108	LABORATORY	All Other	854,482	6.3
	(Unassigned)	DDC	594,140	4.4
502	MENTAL HEALTH-IND	CLI+NOSHOW	296,455	2.2
681	VA-PD HOME/COMM HC	All Other	266,957	2.0
651	STATE NH DAYS	All Other	206,801	1.5
323	PRIM CARE/MED	CLI	173,526	1.3
408	OPTOMETRY	CLI+NOSHOW	169,575	1.2
407	OPHTHALMOLOGY	CLI+NOSHOW	152,636	1.1
180	DENTAL	UTIL	147,875	1.1
	(Unassigned)	All Other	140,923	1.0
411	PODIATRY	CLI+NOSHOW	137,438	1.0
304	DERMATOLOGY	CLI+NOSHOW	109,492	0.8
105	X-RAY	UTIL	108,354	0.8
414	UROLOGY	CLI+NOSHOW	106,194	0.8
560	SUBST ABUSE-GRP	CLI+NOSHOW	102,281	0.8
	All Other ¹⁶	Multiple (not applicable)	4,069,243	29.9
	Total DSS-only records		13,614,689	100

¹⁴ Includes the 46,417 records that were under the NPCD flag and only in DSS as well as DSS records under all other flags.

¹⁵ ‘All other’ is a group identified by excluding records marked by the seven flags.

¹⁶ ‘All other’ represents that group of clinic stops, each of which was less than 100,000 in frequency.

Table 5.18B Clinic stops with more than 100,000 records (stratified by category) among DSS-only non-matched outpatient records FY2004 (includes low cost data only)

CL	CL Label	Category	Frequency	Percent of DSS-only records¹⁷
538	PSYCHOLOGICAL TESTING	All Other	2,099,775	56.2
108	LABORATORY	UTIL	112,564	3.0
	All Other	Multiple (not applicable)	1,525,697	40.8
	Total DSS-only records		3,738,036	100

¹⁷ Includes the 16,058 records that were under the NPCD flag and only in DSS as well as DSS records under all other flags.

Table 5.18C Clinic stops with more than 100,000 records (stratified by category) among DSS-only non-matched outpatient records FY2004 (includes normal cost and low cost data)

CL	CL Label	Category	Frequency	Percent of DSS-only records ¹⁸
538	PSYCHOLOGICAL TESTING	All Other	3,369,622	19.4
423	PROSTHETICS SVCS	PROS	2,462,602	14.2
323	PRIM CARE/MED	CLI+NOSHOW	1,306,909	7.5
108	LABORATORY	UTIL	1,231,833	7.1
108	LABORATORY	All Other	899,011	5.2
	(Unassigned)	DDC	668,899	3.9
502	MENTAL HEALTH-IND	CLI+NOSHOW	322,018	1.9
681	VA-PD HOME/COMM HC	All Other	283,586	1.6
	(Unassigned)	All Other	233,127	1.3
651	STATE NH DAYS	All Other	212,066	1.2
323	PRIM CARE/MED	CLI	183,920	1.1
408	OPTOMETRY	CLI+NOSHOW	183,612	1.1
407	OPHTHALMOLOGY	CLI+NOSHOW	161,133	0.9
180	DENTAL	UTIL	155,984	0.9
411	PODIATRY	CLI+NOSHOW	148,661	0.9
560	SUBST ABUSE-GRP	CLI+NOSHOW	120,279	0.7
105	X-RAY	UTIL	119,221	0.7
103	TELEPHONE TRIAGE	CLI	119,065	0.7
108	LABORATORY	CLI+NOSHOW	117,796	0.7
414	UROLOGY	CLI+NOSHOW	116,002	0.7
304	DERMATOLOGY	CLI+NOSHOW	114,682	0.7
180	DENTAL	CLI+NOSHOW	101,314	0.6
205	PHYSICAL THERAPY	CLI+NOSHOW	100,209	0.6
	All Other	Multiple (not applicable)	4,617,689	26.6
	Total DSS-only records		17,349,240	100

5.9 Other outpatient services in DSS but not in NPCD

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 the DSS NDEs and medical SAS files at Austin is easier than access to PBM. In June of 2003, the DSS team released its first pharmacy

¹⁸ Includes the 62,475 records that were under the NPCD flag and only in DSS as well as DSS records under all other flags.

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 the Denver Distribution Center, contract extended care 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. Researchers should first ensure that 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.

5.10 Pseudo social security numbers

Some records that were generated 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 social security number (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., 000-xx-xxxx). Researchers should exclude these pseudo SSNs from their studies. We listed the scrambled pseudo SSNs and their associated cost statistics in Table 5.19. Among the 8,900 pseudo SSN visits, 8,890 were from laboratory (clinic stop=108), 9 from employee health (clinic stop=999), and 1 from x-ray (clinic stop=105). Note that the number of pseudo records was out of 74 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 608-33-1723 would have (in FY2004) a \$6 million cost (a cost which should be excluded). In FY2003, this scrambled pseudo SSN (608-33-1723) had a \$39 million cost. The VA Decision Support Office (DSO) has released the fiscal year (FY) 2007 technical conversion guidelines. These guidelines establish new rules to make sure that only valid Social Security numbers (SSNs) are used in the database.

Table 5.19 Costs of Pseudo SSNs in DSS in FY2004

SCRSSN	N	Mean (\$)	Maximum (\$)	Minimum (\$)	Sum (\$)
297-33-1383	3	863	1,261	69	2,590
333-33-2333	1	34	34	34	34
409-33-6873	6	1,052	1,261	7	6,310
608-33-1723	8,888	712	36,803	-11	6,327,164
627-33-1143	2	1,261	1,261	1,261	2,521

5.11 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. This leaves 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 a 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.

6. Summary

The DSS national data extracts can be linked almost perfectly with the VHA discharge and outpatient data sets after adjustments in database design are made. The results of the comparisons between the DSS discharge database and the Patient Treatment File (PTF) and between the DSS NDE treating specialty file and the PTF bedsection file were similar to that in FY2003. We compared the DSS discharge database to the Patient Treatment File (PTF), which has clinical data on VA hospital discharges. Records were compared using scrambled social security number (SCRSSN), admission date, station number (STA3N), and discharge date. All but 0.02 percent of the records could be matched. Differences remain in three areas: the bedsection admission and discharge dates in the treating specialty file, and cost outliers.

6.1 Bedsection admission and discharge dates

We compared the DSS NDE treating specialty file and the Patient Treatment File (PTF) bedsection file. We identified records that had the same scrambled social security number (SCRSSN), station number (STA3N), inpatient bedsection admission date, discharge date, and bedsection number in the two files. We found that 73.3% of the DSS bedsection stays matched exactly by those five variables with the PTF database in FY2004. In FY2003, 73.2% were found. These files have different rules for setting bedsection admission and discharge dates, however. Admission or discharge dates sometimes differ by one day; patients admitted in a prior fiscal year were sometimes assigned the first day of the fiscal year as their admission date. When we adjusted for these differences, the comparison attained was very high, as the PTF files included 97.8% of the stays in the DSS treating specialty file. The DSS treating specialty file included 99.7% of the stays recorded in the acute care (PB) PTF bedsection file and 97.4% of the stays in the extended care (XB) PTF bedsection file.

6.2 Missed outpatient records in DSS

We studied FY2004 data and found that 9% of the records in the outpatient National Patient Care Database (NPCD) did not have any care in the DSS normal cost outpatient visits extract. However, when low cost data was added, 99.4% of all NPCD records that were expected to be in DSS were found there. Compared with FY2003, we attained far greater concordance between these files in FY2004 by including low cost DSS encounters in our comparison. The significance of this finding is that by including low cost DSS encounters in the FY2004 comparison between the DSS and NPCD databases, outpatient utilization thought to be missing from DSS was found, and almost all outpatient care was found to be reported in DSS.

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 FY2003 and FY2004, DSS allocated about 21% of the outpatient cost to those services or items that were not in the NPCD. Second, 99.4% of all FY2004 NPCD records that were expected to be in DSS were found there. The DSS Technical Support Office provided a list of reasons for the 0.6% of the records that are only in NPCD (see appendix).

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 (4.9 million patients in FY2004) 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 P. *Reconciliation of DSS Encounter-Level National Data Extracts and the VA National Patient Care Database FY2001-2002*. Menlo Park, CA: VA Health Economics Resource Center; December 2003. Technical Report No. 9.
3. Phibbs C, Yu W, Barnett PG. *Research Guide to Decision Support System National Cost Extracts 1998 - 2004*. Menlo Park: VA HSR&D Health Economics Resource Center; June 2005.

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, and therefore 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: Until FY03, records associated with telephone stops in most cases were not tracked in DSS. 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% or more of all primary care encounters via telephone care. (NB: Beginning in FY03, DSS will collect all telephone stops in NDE.)
3. OBS: Observation encounters are reported in DSS as outpatient encounters stop code 290-296 when NPCD reports these as inpatient 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 for 1-6% of all VAMC mismatches in FY00 in the Seattle HSR&D Study.
6. DUPLICATES ON SAME DAY 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. This results 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 systems to any further revisions. 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 in each fiscal year for periods before September will not be costed.) Starting in FY2004, the DSS NDE has a separate file for outpatient encounters that are assigned low cost by DSS (“low cost” data refer to those

outpatient encounters that are either not assigned costs or assigned costs between -\$1 and \$1). This has improved the correspondence between the DSS and NPCD databases.

11. MEANS SCREEN: Some records from DSS feeds (CLI and UTIL) are present in the DSS national outpatient (OPC) 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.
12. MEANS: DSS collects 3.5 – 6.7% more records than NPCD. The AAC-edit for current means tests removes these from NPCD but DSS collects them via CLI Flag = Y. (See above #9 for the records that never get to NPCD.)
13. REF LAB: Reference Lab patients are reported in DSS as SSN 000123456 (stop code 108). Some may or may not be reported in NPCD by actual SSN.
14. ECS: Some VAMCs did not set up their ECS with the correct stop code, but rather reported the stop code as “ECS.” These patient encounters have the correct stop code in NPCD, but in DSS extracts are found under stop code = ECS.
15. UNK: Some stop codes are UNKNOWN in DSS (usually from non-OR surgical procedures incorrectly set up to feed DSS the correct stop code). These encounters have a 400 series stop code in NPCD, but are found under stop code UNK in DSS.
16. TELEHOME BUDDY: From the second half of FY03, Telehome Monitoring (by non-video monitoring such as Telehome Buddies using web enabling), 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 stop code 429 on the 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 in DSS as NPCD flag – Yes. The actual costs data is found in 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 which are entered into DSS as HCPC costs to a clinic with stop code DDC. All these supply-distribution per SSN encounters are filed in DSS as stop code DDC.
20. PTSD & HOMELESS: Effective in FY02, DSS added records from the PTSD National Mental Health file. In FY03, DSS added records from the National Mental Health file (a) for homeless patients identified in a homeless survey; and (b) for homeless patients who have been discharged from Homeless Rehabilitation programs. None of these cost-free encounters are found in NPCD, but all are found in DSS NDE in FY03 follow-up.