



Guidebook

**HERC's DSS Discharge Dataset with Subtotals for Inpatient
Categories of Care, Fiscal Year 2009**

3rd Edition

Jennifer Y. Scott and Todd H. Wagner

September 2010

**HERC's DSS Discharge Dataset with Subtotals for Inpatient Categories of Care.
Guidebook.**

Health Economics Resource Center (HERC)
VA Palo Alto Health Care System
795 Willow Road (152 MPD)
Menlo Park, CA 94025
650-617-2630
650-617-2639 (fax)
herc@va.gov

Suggested citation: Scott JY, Wagner TH. HERC's DSS Discharge Dataset with Subtotals for Inpatient Categories of Care, Fiscal Year 2009. Guidebook. Menlo Park, CA. VA Palo Alto, Health Economics Resource Center; 2010.

Acknowledgements: We would like to thank Decision Support Office (DSO) for providing information and suggestions toward the creation of this guidebook. We would also like to acknowledge the financial support from the VA Health Services Research and Development (HSR&D).

Contents

Terms	v
1. Overview	1
2. Merging the Discharge (DISCH) and Treating Specialty (TRT) Files	2
2.1 Categories of Care	2
2.2 Adding Category Costs and Lengths of Stay to the DISCH	3
3. Methods of Reconciliation	6
3.1 Cost Reconciliation	6
3.2 Length of Stay Reconciliation	8
4. Important Notes for Using the HERC Discharge Dataset with Subtotals	10
4.1 Access to the HERC Discharge Dataset	10
4.2 New Variables Added to DSS DISCH File	10
4.3 No Reassignment of Missing Values	11

Tables

Table 1: Assignment of Treating Specialties to Categories of Care, FY 2009 Update	2
Table 2: Frequency for Cost Method Variable, CST_METH, by Fiscal Year	6
Table 3: National Daily Average Cost for Each Category of Care by Fiscal Year	7
Table 4: Frequency of Length of Stay Method Variable, LOS_METH, by Fiscal Year	8
Table 5: HERC Discharge Files by Fiscal Year	10

Terms

AAC	Austin Automation Center
AITC	Austin Information Technology Center
DISCH	Discharge National Data Extract
DSO	Decision Support Office
DSS	Decision Support System
FY	Fiscal Year
HERC	Health Economics Resource Center
ICU	Intensive Care Unit
LOS	Length of stay
NDE	National Data Extract
PRRTP	Psychosocial Residential Rehabilitation Treatment Program
PTF	Patient Treatment File
SCI	Spinal Cord Injury
TRT	Treating Specialty National Data Extract
VA	U.S. Department of Veterans Affairs
VHA	Veterans Health Administration

1. Overview

Decision Support System (DSS) maintains National Data Extracts (NDE) that track cost and utilization for care provided by the U.S. Department of Veterans Affairs (VA) medical centers. Researchers can access these files, which are stored as SAS datasets on the VA mainframe in Austin, Texas.

The DSS Discharge (DISCH) NDE includes information on the entire span of an inpatient hospitalization. It provides the discharge bed section (DBEDSECT), but does not have detailed information on other treating specialties during an inpatient stay. A bed section, also known as the treating specialty, is a two-digit code developed by VA to characterize the type of hospital care.

If a researcher is interested in information on specific treating specialty segments of an inpatient stay, then the researcher must extract the information from the DSS Treating Specialty (TRT) NDE. To expedite this process, we created a new dataset beginning fiscal year (FY) 2007 that is identical to the DSS DISCH NDE with the exception of additional fields containing cost and length of stay subtotals for each inpatient category of care (e.g., acute medicine, psychiatry, nursing home, etc.). The categories of care represent our groupings of common bed sections as reported in the TRT NDE.

This guidebook describes how we prepared the HERC Discharge dataset with cost and length of stay (LOS) subtotals for each inpatient category of care. [Chapter 2](#) describes the methods we used to merge the DSS DISCH and TRT. In our comparison of the two files, we found that the vast majority of the records had equivalent total costs and lengths of stay. There were few cases where this was not true. [Chapter 3](#) describes our methods for dealing with the small number of records with cost and length of stay differences between the DISCH and TRT, while [Chapter 4](#) discusses important information regarding the use of the new Discharge dataset.

For more information on these DSS NDEs or other available NDEs, see [National Data Extract \(NDE\) Technical Guide, October 2009](#) or [Research Guide to Decision Support System National Cost Extracts](#). **Note:** These links direct you to internal VA websites and are not available to the public.

2. Merging the Discharge (DISCH) and Treating Specialty (TRT) Files

The DSS DISCH file includes expenses associated with a completed inpatient encounter and is reported in the fiscal year of discharge. Each record represents a unique hospital stay. Consequently, the DISCH may include information from multiple fiscal years. The TRT, on the other hand, corresponds to a patient's duration in a treating specialty segment. Each segment is defined either by a change in treating specialty, or for each complete month if no change in treating specialty took place. For example, a veteran who is hospitalized in neurology from April through May will have two records. Therefore, multiple TRT records can exist for a given admission because patients move to different treating specialties during their hospitalization, or they stay past the end of the month in the same treating specialty. TRT has information on the hospital care provided in the current year, regardless of whether the patient was discharged. Information on stays that span multiple fiscal years is in multiple TRT files.

2.1 Categories of Care

Thirteen categories of care were created based on the treating specialty variable, TRTSP, in the TRT dataset. Each category represents common groupings of bed sections, but is not completely consistent with the HERC Inpatient Average Cost categories. The HERC Discharge has a separate category for both intensive care unit (ICU) and unidentified treating specialties (see Table 1 below). For a list of treating specialties and their descriptions, please visit http://vaww.aac.va.gov/npcd/PTF_TS_CODES.htm. **Note:** This is an internal VA website and is not available to the public.

Table 1: Assignment of Treating Specialties to Categories of Care, FY 2009 Update

Inpatient Category of Care	Treating Specialty Number(s)
0: Acute Medicine	Missing, 1-11, 14-19, 24, 30, 31, 34, 83, 1E, 1F, 1H, 1J
1: Rehabilitation	20, 35, 41, 82, 1D, 1N
2: Blind Rehabilitation	21, 36
3: Spinal Cord Injury (SCI)	22, 23
4: Surgery	48-62, 65, 78, 97, 1G
5: Psychiatry	25, 26, 28, 29, 33, 38, 39, 70, 71, 75-77, 79, 89, 91-94, 1K, 1L
6: Substance Abuse	27, 72-74, 84, 90
7: Intermediate Medicine	32, 40
8: Domiciliary	37, 85-88
9: Nursing Home	42-47, 64, 66-69, 80, 81, 95, 96, 1A, 1B, 1C
10: PR RTP*	25-29, 38, 39
11: ICU	12, 13, 63
999: Unidentified	Unidentified treating specialty

* Only for stations 459, 463, 501, 504, 515, 516, 518, 523, 528, 541, 546, 549, 554, 555, 556, 561, 568, 573, 586, 589, 590, 595, 598, 620, 622, 631, 632, 635, 637, 640, 645, 653, 656, 658, 662, 663, 666, 676, 678, 687, and 689.

2.1.1 Change to Treating Specialty Variable Type **NEW!**

Beginning FY 2007, the treating specialty variable, TRTSP, was converted from numeric to character variable type. This change was made to accommodate new alphanumeric treating specialty codes (refer to Table 1). TRTSP_C, the former character variable type for treating specialties, has been dropped because it is no longer needed.

2.1.2 Assumption Regarding Missing Treating Specialties

In FY 2008, a diminutive amount of TRT records (approximately .3% of all TRT records from the same fiscal year) had missing treating specialties representing valid costs and utilization. We were unable to ascertain the treating specialty by linking the TRT with the Patient Treatment File (PTF) Bed Section file. However, we required a technique that could easily be replicated for future versions of the HERC Discharge dataset. We assigned missing treating specialties to Acute Medicine since over 50% of inpatient care provided by VA is in this category. And because we are dealing with so few observations, it would not make a large impact on subtotal summarizations.

2.2 Adding Category Costs and Lengths of Stay to the DISCH

Costs and lengths of stay were calculated for each category and summed across unique combinations of scrambled social security number (SCRSSN), station (STA3N), admission date (ADMITDAY), and encounter number (ENC_NUM). These variables were used to identify a unique hospital stay. When we combined the DISCH file with the summarized TRT dataset by these key variables, nearly 100% of the records merged perfectly. The remaining few unmatched DISCH records were found to have zero costs and missing encounter numbers even though the lengths of stay were greater than zero. Because no ENC_NUM value was available for these DISCH records that could not be matched to a unique hospitalization in TRT, we used a SAS DATA step to perform another merge on these few records. The BY variables we used were SCRSSN, STA3N, and ADMITDAY (excluding the ENC_NUM variable that we used in the first merge at the beginning of this section). After this stage, no DISCH records remained unmatched to a summarized TRT record which included subtotals for categories of care.

2.2.1 Changes to Inpatient Variables in FY 2008

In FY 2008, the total DISCH cost variable, DCST_TOT, along with the total TRT cost and length of stay variables, TCST_TOT and TRT_LOS, respectively, were dropped. DSO has already corrected this in the FY 2009 core NDEs. To calculate the total costs in FY 2008, we summed all the fixed direct, fixed indirect, and variable direct subtotals. We derived the length of stay in TRT by using the following algorithm if TRTIN contained a non-missing value: $\max(\text{TRTOUT} - \text{TRTIN}, 1)$. In other words, we calculated the length of stay by subtracting the day the patient entered the treating specialty (TRTIN) from the day the patient exited the treating specialty (TRTOUT).

2.2.2 Changes to Inpatient Variables in FY 2009 **NEW!**

In FY 2009, the DISCH and TRT total cost variables were included in the NDEs. However, the length of stay variable, now called LOS instead of TRT_LOS, has missing values for all records. Therefore, we calculate the length of stay using the following formula that we employ several times throughout the construction of the HERC Discharge file: $\max(\text{TRTOUT} - \text{TRTIN}, 1)$. (This method is described in more detail in the next section.) However, we found some cases where the TRTIN variable, date of entry into treating specialty, was missing. To reconcile these records, we either took the first day of the TRTOUT month or the admission date, ADMITDAY, whichever occurred second. There were no cases of missing TRTOUT dates. As described at the beginning of this chapter, each TRT record is represented either by a change in treating specialty, or for each complete month if no change in treating specialty took place. Therefore, if a patient was admitted to a VA hospital on 7/22/2009 and exited the treating specialty on 7/31/2009, then the missing entry date into the treating specialty would be coded to reflect 7/22/2009 because it occurs after the first of July. On the other hand, if a patient was admitted on 3/27/2008 and had an exit date recorded as 4/30/2008, then the new entry date would be 4/1/2008 because it occurs after the admission date. If this patient did not change treating specialties, the dates tell us there should be at least two TRT records for this encounter, one from 3/27/2008 to 3/31/2008 and one from 4/1/2008 to 4/30/2008. It is possible that the patient's stay extended past 4/30/2008.

2.2.3 Definitions of Key Variables for DATA Step Merge

The DISCH and TRT contain information about VA hospitalizations. The variables below uniquely identify a stay and were used to link the DISCH and summarized TRT files using a MERGE statement in the SAS programming language.

- **Scrambled social security number (SCRSSN).** A unique patient identifier formatted to resemble a social security number.
- **Station (STA3N).** A three-digit code representing the VA medical facility where the patient received care.
- **Admission date (ADMITDAY).** The date of admission into a VA medical facility. This is different from the TRTIN variable in the TRT NDE, which represents the date of entry to a treating specialty.
- **Encounter number (ENC_NUM).** An encounter number is a character string consisting of admission date (YYMMDD), the letter "I", and possibly a sequence number depending on the number of admissions on the same day. A second admission on the same day creates a suffix of "1," a third admission on the same day creates a suffix of "2," and so on. For example, if a patient was admitted to a VA hospital on 1/20/2007, discharged and moved to a different facility on the same day, there would be two encounter numbers for the two separate admissions, 070120I and 070120I1, respectively. If the patient did not move between facilities and only had one admission, then only 070120I would be listed.

2.2.3 Using TRT Files from Prior Fiscal Years

The DISCH includes inpatient hospital stays that ended in a fiscal year. The TRT provides information on hospital care that was provided in a specific fiscal year. When a stay crosses multiple fiscal years, it is characterized by a single DISCH record in the year of discharge and many TRT records in two or more fiscal years. Therefore, we aggregated TRT datasets from FY 2000 (earliest available) onward to account for information not found in the TRT file of the year of discharge. However, TRT files from certain fiscal years do not contain the variables needed to merge with the DISCH file. A list of variables with missing information is presented below.

- **Encounter numbers.** Valid encounter numbers, represented by the variable ENC_NUM, only exist for FY 2004 and FY 2006 onward. We defined “valid” to mean those with a 7 or 8 character string consisting of a date (YYMMDD), the letter “T”, and possibly a sequence number. We found that all values in the ENC_NUM field were zero in FY 2003.
- **Length of stay.** We calculated treating specialty segment lengths of stay (TRT_LOS) for FY 2000-2004, 2008, and 2009 because no equivalent variable was available for those years. We used the following algorithm to calculate length of stay: $\max(\text{TRTOUT} - \text{TRTIN}, 1)$ ¹. In other words, if the exit date minus the entry date equals zero, then TRT_LOS would equal 1 since an inpatient stay in a treating specialty is at least one day.
- **Discharge date.** Discharge date (DISDAY) was not included in TRT until FY 2007. We did not use this variable to merge with the DISCH file.

After merging the DISCH and summarized TRT with subtotals, we compared the totals costs and lengths of stay from each file. Consistently over the years, almost 100% of all records had equal or nearly equal costs, while almost 99% had equal or nearly equal lengths of stay. [Chapter 3](#) discusses in detail the methods we used to reconcile the small number of differences.

¹ Where TRTOUT = exit date from treating specialty, TRTIN = entry date into treating specialty, and AGGABS = aggregate absent days.

3. Methods of Reconciliation

3.1 Cost Reconciliation

Information on each hospital stay from the DISCH and TRT was combined into a single file that had complete data from both datasets. The DISCH identified the total cost for the entire stay, as well as total length of stay. The TRT file identified the cost of care and length of stay in each hospital bedsection. We compared the cost reported in the two files to check that the merge was correct and that the data sources were consistent. Nearly all stays (over 99.9%) between the DISCH and TRT had equal costs, or cost differences of less than \$100 which we term as “equal”. The following sections discuss in detail the methods we used to reconcile these records that represent less than .1% of all discharges in the fiscal year.

3.1.1 Cost Method Variable, CST_METH

CST_METH is a categorical variable we created to flag records based on the type of reconciliation used when there were varying cost differences. Most of the records were assigned a value of “0” because no reconciliation was required. In other words, the DISCH cost and the total TRT cost (sum of category costs) differed by less than \$100 (nearly all differed by less than \$10). Records reconciled using a proportion-based method described in section 3.1.2 were assigned “1”. A value of “2” was given to records we reconciled using national average daily costs as described in section 3.1.3. Finally, a value of “3” was assigned to a handful of records with incomplete data. The subtotals for these records reflect the information we have from FY 2000 (earliest available TRT NDE) onward. Because these admissions were prior to FY 2000, you will find that the total DISCH cost is always greater than the sum of costs for all categories of care. Please refer to Table 2 below for the number of records assigned to each value of the cost method variable, CST_METH.

Table 2: Frequency for Cost Method Variable, CST_METH, by Fiscal Year

CST_METH	FY 2007	FY 2008	FY 2009
0: No reconciliation	641,353	658,102	676,459
1: Proportion-based method	299	0	0
2: National average daily costs	118	76	44
3: Incomplete	109	77	73
<i>Total</i>	641,879	658,255	676,576

3.1.2 Proportion-Based Method for Cost Reconciliation

In FY 2007, we used a proportions-based method to estimate subtotals for records where patients received care at stations 618 (Minneapolis, MN) or 679 (Tuscaloosa, AL). Our findings showed that records from these stations not only had large cost differences, but that the stays also began and ended in FY 2007. According to DSO, there was a costing problem just before the end-of-year NDE’s were created that only affected these two stations. The DISCH dataset provides the correct total costs. To reconcile the differences among these few records, we calculated the

proportion of each category cost from the total TRT cost and then applied this proportion to the total DISCH cost to generate new subtotals. All records resolved with this approach were flagged in CST_METH with “1”.

Example of Using Proportion-Based Method

A TRT record consists of three category costs: \$500 in category 1, \$750 in category 4, and \$250 in category 5. If the DISCH cost was \$2,000, we would multiply \$2,000 by 33.3% ($500/1,500$) to get a new category 1 cost of \$666.67, 50% ($750/1,500$) to get a new category 4 cost of \$1,000, and 16.7% ($250/1,500$) to get a new category 5 cost of \$333.33. The sum of these new category costs now equal the total DISCH cost of \$2,000.

Starting FY 2008 onward, the proportion-based method was not employed because there were no records with large cost differences when both admission and discharge occurred in the same fiscal year. Therefore, in [Table 2](#), CST_METH=1 has a zero count after FY 2007.

3.1.3 Applying National Average Daily Costs

Of the records with cost differences greater than \$100 and admission dates in FY 2000 through last fiscal year, we could not determine whether the DISCH cost or summarized TRT cost was correct. Therefore, our resolution was to apply national average daily costs (see Table 3). These costs were generated from the TRT NDE. The only modification made to the original dataset was the inclusion of category of care costs and lengths of stay. Once the national daily average costs were calculated, we multiplied these values by the length of stay² for each category at the stay-level. The results were then adjusted by the ratio between the total DISCH cost and the new total TRT cost. Records reconciled in this manner were flagged with a number “2” in the cost method variable, CST_METH.

Table 3: National Daily Average Cost for Each Category of Care by Fiscal Year

Inpatient Category of Care	FY 2007	FY 2008	FY 2009
0: Acute Medicine	\$1,831	\$2,060	\$2,214
1: Rehabilitation	\$1,480	\$1,770	\$1,972
2: Blind Rehabilitation	\$1,058	\$1,198	\$1,342
3: Spinal Cord Injury	\$1,372	\$1,551	\$1,724
4: Surgery	\$2,757	\$3,160	\$3,428
5: Psychiatry	\$994	\$1,105	\$1,235
6: Substance Abuse	\$547	\$610	\$622
7: Intermediate Medicine	\$1,364	\$1,459	\$1,579
8: Domiciliary	\$270	\$308	\$337
9: Nursing Home	\$713	\$814	\$913
10: PR RTP	\$400	\$458	\$502
11: ICU	\$4,219	\$4,703	\$4,954
999: Unidentified	\$2,023	\$104	\$188

² Note that these averages are based on unadjusted lengths of stay.

3.1.4 Records with Incomplete Data

Records with admissions prior to FY 2000 were flagged with a “3” in CST_METH. Category costs associated with these records are incomplete and do not reflect the entire stay because data only exist for FY 2000 onward. Note these records represent hospitalizations in VA nursing homes spanning almost a decade on average. Therefore, you will find significant cost differences when compared to the total DISCH cost for the stay.

3.2 Length of Stay Reconciliation

Similar to the cost comparison, we examined the length of stay differences between the DISCH and summarized TRT file. Over the years, almost 99% of the records consistently did not require reconciliation because the lengths of stay were equal between the two datasets, where we defined “equal” to be a difference of 5 days or less.

3.2.1 Length of Stay Method Variable, LOS_METH

LOS_METH is a categorical variable we created to flag records based on the type of reconciliation used for length of stay differences between the DISCH and summarized TRT file. Almost all the records were assigned a value of “0” since no reconciliation was needed. A value of “1” was given to records where the DISCH length of stay was missing and, therefore, was calculated based on the formula described in the next section. We reassigned length of stay to about 1% of the records. These observations were assigned a value of “2” in the LOS_METH variable. Lastly, records reconciled with the proportion-based method received a “3”.

Table 4: Frequency of Length of Stay Method Variable, LOS_METH, by Fiscal Year

LOS_METH	FY 2007	FY 2008	FY 2009
0: No reconciliation	634,429	648,838	666,771
1: Calculated length of stay	20	26	29
2: Reassign length of stay	6,845	7,048	7,987
3: Proportion-based method	585	2,343	1,789
<i>Total</i>	641,879	658,255	676,576

3.2.2 Records with Missing DISCH Lengths of Stay

There are usually a handful of records with missing values in the DISCH length of stay variable, LOS. In FY 2007, all utilization for these records occurred in acute medicine (category 0). Because these records only required one subtotal, we calculated the length of stay and assigned this value to the length of stay variable for acute medicine, LOS0. The length of stay calculation was based on the following formula: $\max(\text{DISDAY} - \text{ADMITDAY} - \text{AGGABS}, 1)$ ³. In other words, if the discharge date minus the admission date minus the aggregate absent days equaled zero, then LOS0 would equal 1 since an inpatient stay is at least one day. We gave these records a

³ Where DISDAY = discharge date from VA medical facility, ADMITDAY = admission date to VA medical facility, and AGGABS = aggregate absent days.

flag of “1” in LOS_METH. Missing values in the DISCH variable, LOS, were not changed to reflect the new length of stay calculation. The reason we did not assign LOS0 the TRT length of stay total was because these values appeared to be inaccurate.

In FY 2008, some of these encounters with missing DISCH length of stay had utilization in more than one category of care (i.e., in both acute medicine and unidentified categories). We used the method described earlier in this section to reconcile records with utilization occurring in only one category (acute medicine or unidentified). There were only two records with utilization in both acute medicine and unidentified categories. Therefore, the approach described above could not be used to reconcile these records. Instead, we employed the method described in Section 3.2.4. If this occurred in future years, the same method of reconciliation was employed.

3.2.3 Utilization in Only One Category of Care

There were other records with utilization occurring in only one category. These records; however, had non-missing DISCH lengths of stay. We assigned the DISCH length of stay to the category where utilization occurred. These records were given a value of “2” in LOS_METH. We did not assign the TRT length of stay because the values appeared to be inaccurate based on our calculations. We also wanted to refrain from changing the original data in the DSS DISCH NDE.

3.2.4 Proportion-Based Method for Reconciling Length of Stay Differences

The remaining records (less than .1%) were reconciled using the same proportion-based method employed in section 3.1.2. Since utilization occurred across multiple categories, we could not simply reassign the DISCH length of stay. Using the proportion-based method, we rounded the lengths of stay to the nearest whole day and flagged them with the number “3” in the LOS_METH variable.

4. Important Notes for Using the HERC Discharge Dataset with Subtotals

The HERC Discharge dataset is functionally identical to the original DSS DISCH file upon which it is based. No modifications have been made to the original data. We have only included new variables corresponding to cost and length of stay subtotals, as well as two flag variables.

4.1 Access to the HERC Discharge Dataset

The HERC Discharge dataset has been stored as a SAS file at Austin Information Technology Center (AITC)⁴ and can be found in the RMTPRD.HERC.SAS directory. We have listed the file names, as well as information on the number of records and variables, by fiscal year in Table 5 below.

Table 5: HERC Discharge Files by Fiscal Year

Fiscal Year	File Name	Number of Records	Number of Variables
2007	RMTPRD.HERC.DSS.SAS.HDISCH07	641,879	103
2008	RMTPRD.HERC.DSS.SAS.HDISCH08	658,255	96
2009	RMTPRD.HERC.DSS.SAS.HDISCH09	676,576	102

4.1.1 Required Task Codes

Users of the HERC Discharge dataset will need permission to access DSS national extracts. For more information on obtaining these permissions, including the required task code, please contact us at herc@va.gov. You can also refer to [HERC's Research Guide to Decision Support System National Cost Extracts](#). **Note:** This is an internal VA website and is not available to the public.

4.2 New Variables Added to DSS DISCH File

HERC adds at least 28 new variables to the DSS DISCH NDE (created by VHA DSO) every fiscal year. These correspond to the 13 cost variables and 13 lengths of stay variables for each for each category of care, as well as the two flag variables indicating which method was used to reconcile differences. Subsequent variables are added on an as-need basis. For instance, in FY 2008, total costs were excluded from both the DSS DISCH and TRT NDEs. HERC calculated these variables and included them in the final dataset.

Special attention should be given to the cost method (CST_METH) and length of stay method (LOS_METH) variables. Depending on the type of reconciliation we employed, researchers may want to exclude certain records from their analyses. For instance, an analyst may want to query

⁴ Formerly known as Austin Automation Center (AAC).

records where `CST_METH=0` and `LOS_METH IN (0,2)`⁵ because no modifications were made to the cost and length of stay categories.

Researchers familiar with the HERC Average Cost datasets may know the distinction between local and national costs. DSS provides local costs in their datasets and; therefore, the subtotals we calculated in the HERC DISCH dataset should also be considered local costs. For information regarding local and national cost estimates, please refer to [*HERC's Average Cost Datasets for VA Inpatient Care FY1998 - FY2009*](#). **Note:** This is an internal VA website and is not available to the public.

4.3 No Reassignment of Missing Values

Costs and lengths of stay have missing values in categories with no utilization. However, if there is utilization but no cost, then “\$0.00” will appear in the category where the patient received care, while the other categories for that record will have missing values.

We did not assign zero to missing costs or lengths of stay for the following reasons:

1. There are some missing values for length of stay, LOS, in the DISCH file. As a result, we wanted to be consistent throughout the whole dataset.
2. It is easy to zero-fill missing values, but difficult to remove zero-fills.
3. Most importantly, we did not want to affect summarizations made by researchers using our dataset. For example, if we assigned zeros, then the average cost for a category of care would be lower than expected. If an analyst uses the MEANS procedure in SAS to summarize costs or lengths of stays, for instance, missing values would be excluded from such calculations.

⁵ This code uses the IN operator in the SAS programming language.