

## **Data and Methods**

This describes data sources and methodological approaches used in the research. First, the researchers described the data sources and the creation of the 2001 SNF analysis file. Second, they addressed the selection of units of observation, and described their approach for dealing with timing inconsistencies between claims data on charges and MDS data on patient characteristics. Third, they briefly described the dependent and explanatory variables that were examined.

### **A. Data Sources**

The principal data source for this study is the 2001 Data Analysis PRO (DataPRO) file of Medicare SNF stays that are linked with minimum data set (MDS) assessments, as well as information merged from the qualifying, or prior, acute care hospital stays. DataPRO was created to facilitate a medical review process of Medicare SNF residents, following the implementation of the SNF PPS, and contains nationally representative information on Medicare SNF stays. Approximately 2 million such stays occur each year. The researchers enhanced the DataPRO stay records with additional information from Medicare claims including charges for specific types of services (e.g., respiratory services, prescription drugs) used during the SNF and the prior hospital stays.

The MDS is part of an overall nursing home resident assessment system, required by the Nursing Home Reform Act of OBRA 1987, which was developed to improve the health and quality of life of nursing home residents. The MDS consists of over 300 questions grouped into 18 domains, such as patient diagnosis, cognitive status, functional status, and nutritional status. Of the roughly 300 items on the full assessment form, about 125 are used for PPS payment purposes. MDS assessments for SNF residents are made on a specified schedule approximately 5, 14, 30, 60, and 90 days from the start of the Medicare stay. In DataPRO, all MDS assessments for a resident are matched to each SNF stay, with selected items pulled for the file.

Medicare SNF and hospital providers submit claims for reimbursement for Medicare-covered services. Information on these claims includes periods of service, types of procedures furnished, primary and secondary diagnoses of the resident, and the institution's charges for services provided. Claims are submitted to Medicare fiscal intermediaries, contractors that conduct edits, review for appropriateness, and reimburse the provider according to Medicare eligibility, coverage and payment rules. Claims from the Medicare intermediaries are ultimately sent to CMS.

Medicare participating SNFs submit cost reports annually to fiscal intermediaries. Among other things, these reports itemize Medicare-related costs for routine and ancillary services, and cost of capital. The researchers used SNF cost report data to derive routine costs and ancillary service cost-to-charge ratios (CCRs) that were used to convert ancillary service charges from claims to estimated costs for those services. Because charges for routine services (e.g., nursing, accommodations) are not generally differentiated on the claims for residents in the same facility, they assigned the per diem routine costs of a resident based on each SNF's reported routine service costs in the cost reports.

## **B. Construction of the 2001 Analysis files**

The analysis file for this study contained Medicare SNF stays with information on estimated Medicare costs for different service components. Given the need to estimate ancillary costs from claims by applying CCRs from cost reports to claims, the eventual analysis file had to contain 2001 SNF stays that: (a) had cleanly matched MDS information on resident characteristics, (b) came from facilities for which a 2001 cost report was available, (c) had consistent claims information on SNF and hospital services not captured by DataPRO, and (d) had internally consistent information from the multiple sources of data.

The researchers used the DataPRO SNF and prior hospital stays as the starting point for creating the analysis file. They examined the stays to identify potential integrity problems and non-Medicare coverage and found anomalies, such as overlapping claims records, zero covered days, and missing prior hospital stay information. They then examined the MDS assessments associated with each stay and found other anomalies, such as absence of assessments, irregular patterns of assessments, and mismatches between MDS assessments in DataPRO and MDS assessments from the MDS core data files. Next, they examined consistency of information between DataPRO stays and SNF stays that they created from raw claims data (which would be the source of service use information not originally collected by DataPRO). Medicare stays that were found to have inconsistent information were excluded from the analysis file.

The next step was to determine the intersection between the “clean” Medicare SNF stays and the SNF cost reports on record. SNF stays without corresponding cost report data—which could not have the claims amounts converted to estimated costs—were also excluded from the analysis files.

## C. The Number of Cases Remaining After Specific Initial Edits

<i>Reasons for Exclusion of Stays</i>	<i>Remaining Numbers</i>
Total SNF stays in 2001 DataPRO	2,114,797
1. After exclusions for “Integrity Problems and Non-Medicare coverage”	1,900,036
• Overlapping SNF stays, fragmented SNF stays, overlapping claims records within stays, no Medicare payment, no covered days (3 percent)	
• Overlapping qualifying hospital stays, no qualifying hospital stay, overlapping claims records for the qualifying hospital stay, no Medicare payments (2.8 percent)	
• Non-PPS stay (5.8 percent)	
• Swing Bed stay (5.7 percent)	
2. After exclusions because of unavailability of MDS assessments	1,778,059
• No MDS assessments (4.8 percent)	
• No 5-day MDS assessment (.8 percent)	
• Missing complete sequence of assessments (e.g., a stay with 5-day and 30-day assessments, but no 14-day assessment) (.7 percent)	
• Mismatch between SNF stays’ assessments and base MDS data set (.05 percent)	
3. After exclusions because of mismatch of the DataPRO stays and Urban Institute stays <sup>1</sup>	1,768,761
• Non-match of stays (SNF stays or hospital stays)	
• Stays with different SNF covered days	
• Stay with different hospital covered days	
4. After matching with cost report information and wage-index file (13,718 SNF providers matched)	1,722,987
5. After other exclusions	1,709,736
• Length of stay is not equal to Medicare covered days (.8 percent)	
• Stays that have the same assigned MDS assessments (80 stays)	

## D. Samples and “Base Case” Facility and Stay Data Exclusions

The original analysis file contained 1.7 million Medicare SNF stays in 2001, and they created approximately 600 variables for each stay. Due to the enormous amount of data available, the researchers chose to select a 10% random sample of stays for the purpose of developing case-mix classification models, which are referred to as the “10% stay file” or “simple random sample file.”

---

<sup>1</sup> Although covered days are the same on both SNF stays and qualifying hospital stays, there are a few thousand stays where total charges on the DataPRO SNF stays and the UI SNF/Hospital stays do not agree. They decided to use the UI constructed charges and per diem charges.

A second sample was drawn for validation and facility-level analyses. A random 10% sample of facilities was identified and all stays in those facilities were included in the second sample. This file is called the “10% facility/stay file” or the “validation sample.”

A uniform set of additional “base case” exclusion rules for facilities and stays were applied to both samples that are used across all approaches. The various resident classification approaches were compared using the results that obtain for the common base case, though some analyses may have considered alternative rules as well. The base case also entails a common set of statistical procedures discussed in other sections.

### ***Facility data exclusion rules (flags)***

The researchers next identified facilities whose data was likely to be unreliable and should be dropped from an analysis sample. There were three types of problems that were encountered on facility cost reports that led to the creation of flags for specific facilities: 1) unreliable cost data on the cost report, 2) an inability to accurately assign Medicare costs from the total facility costs, and 3) unreliable cost to charge ratios.

#### ***1) Unreliable cost data***

Unreliable cost data was determined based upon unlikely cost breakdowns or extreme cost levels. The researchers flagged facilities that had unacceptable ratios of ancillary to routine costs based upon ratios less than or equal to .005 or ratios greater than 5.0. Inconsistent facilities that had total ancillary costs exceeding the sum of the costs of the ancillary cost components by a significant amount were identified as unreliable. They also flagged four facilities with the following extreme costs: routine cost per day of \$0.009, routine cost per day of \$36,409, routine cost per day of \$0.28, and NTA cost per day of \$8,815.

#### ***2) Unable to assign Medicare costs in a facility***

The researchers needed to assign Medicare costs in a facility in order to generate CCRs for calculating costs for the observations in the file. Facilities that were missing data necessary to assign Medicare costs were flagged. These were facilities missing any of the following data: number of SNF beds, number of Medicare resident days, number of SNF participating unit days, and number of NF unit days.

#### ***3) Unreliable CCRs***

Because of the interest in examining different levels of aggregation for ancillary services, information was “built up” to the desired levels. For example, the sum of department specific costs for NTA services and the sum of department specific charges for NTA services were used to construct the CCR for all NTA services (rather than using the “total” cost and charge fields on the cost reports).

The researchers sought to apply CCRs at disaggregated levels of SNF ancillary services. TAP panel members suggested separating rehabilitation therapy from NTA services. Because a major focus of this study is to develop classification of residents by NTA costs, they explored the possibility of disaggregating NTAs into more distinct

components: (1) drugs, (2) respiratory, and (3) all other NTAs. In the analysis, they examined the CCRs for those 3 components of NTAs, as well as for: (4) total NTAs, (5) rehabilitation therapy, and (6) total ancillary services.

They conducted an analysis to determine which CCRs appeared to be “out of range.” They flagged SNFs with such ratios for any of the service components that they examined. It is important to note that there is no right answer, and this analysis served to help make decision rules. The goal was to be as inclusive as possible, while identifying particular SNFs that were likely to contribute erroneous data for our classification analysis.

They first explored ranges of CCRs that were used by other researchers in Medicare PPS-related research. In research on inpatient rehabilitation facilities, RAND settled on a range of .05-10.0 for departmental CCRs; when values were outside of this range, values were imputed based on like facilities. In earlier research on SNF refinement, Abt Associates employed a range of 0.25-2.1 for ancillary services. In research on acute care hospital CCRs, Newhouse, et al. (1989) employed limits for departmental CCRs that varied by size of hospitals; .01- 100 for hospitals with less than 100 beds and .01-3.0 for hospitals with 100 or more beds.

They conducted detailed analyses of CCR distributions for hospital-based and freestanding SNFs separately, flagging SNFs that had CCRs that appeared to be extreme values. From this analysis, they determined a range of .05 – 30.0 as a reasonable starting point for (a) total ancillary, (b) rehabilitation therapy, (c) total NTA, and (d) drugs. For respiratory and other NTA, they determined that a higher cap would be appropriate. They selected the range of .05 to 100 for these last two components. Freestanding SNFs with CCRs outside these ranges were flagged. They flagged hospital-based SNFs that had service specific CCRs that fell outside of the .05-10.0 range.

The impact of applying these flag rules on number of SNFs and number of associated stays, separately for freestanding and hospital-based SNFs, is relatively minor. For freestanding SNFs, the 112 flagged SNFs and 1,175 associated stays are 0.9% of all freestanding SNFs and 0.9% of all freestanding SNFs’ stays. For hospital-based SNFs, the flagged cases are 0.8% of all hospital-based SNFs and 0.3% of all stays in hospital-based SNFs.

The researchers determined that the “generous” range of CCRs that were applied enabled them to include virtually all of the stays in freestanding and hospital-based SNFs respectively. This high yield led us to consider excluding the small number of stays from the analysis, and not attempt to impute values for them.

A related issue is that some claims are submitted for particular types of services (e.g., respiratory), for which CCRs could not be constructed using a facility’s cost reports. In these cases, they applied the available CCR from the next higher level of service aggregation. For example, if a claim for respiratory therapy is submitted, but the SNF does not have a respiratory CCR, they applied the CCR for total NTA. Similarly, if a claim for rehabilitation therapy is submitted and no rehabilitation therapy CCR exists, they applied the CCR for total ancillary services.

### *Stay data exclusion rules (flags)*

Finally, they eliminated stays with extreme values for total ancillary costs and charges out of concern for the validity of the data. They dropped stays with logged total ancillary costs or charges 3 standard deviations from the logged mean. Most of the stays identified through costs overlapped those identified with extreme charges and so the process eliminated about 2.0% of the remaining observations. The number of stays in each sample before and after the base-case exclusions is given in Table 1.

**Table 1 Number of SNF Stay Observations in the 10% Stay and 10% Facility/Stay Analysis Samples**

	<b>10% Stay File (N)</b>	<b>10% Facility/Stay File (N)</b>
Before exclusions	170,774	177,960
After facility exclusions	167,113	174,263
After stay exclusions	163,738	170,783

### **E. Selecting Units of Observation**

The key limitation in the analysis file data is that ancillary service charges (and, thus, costs) are not reported for specific dates of service, but have to be averaged and applied to the entire SNF stay. Hence, the MDS assessments on resident conditions, which are recorded at specific dates during a stay, cannot be linked directly to specific cost information. Because approximately 60 percent of the SNF stays have multiple MDS assessments, it is important to consider if, and how, information from all the assessments can be employed in the prediction of costs.

The researchers developed a procedure to make efficient use of all of the MDS assessments for each stay. The reasoning behind this approach is that the outcome – per diem cost measured over the entire stay – is a weighted average of the per diem costs measured over the MDS assessment periods, or segments. As a result, it is appropriate to examine cost variation with the weighted average of the resident conditions over the same assessment periods, or segments. This approach is achieved by defining segments for each MDS and weighting them together based on the proportion of time during the stay that is covered by each assessment.

### **F. Cost Variables**

In this report they analyzed stay-level cost data. They examined predictors of major components of SNF services, such as routine, rehabilitation therapy and NTA service costs, as well as total SNF costs. Because a major focus of this study is to predict NTA costs, they also investigated the determinants of 3 components of NTA, notably prescription drugs, respiratory therapy, and other NTA components combined.

Because preliminary analyses indicated that a much higher proportion of the variance in costs per diem could be explained than that of costs per stay, they focused their analysis on costs per diem. They also compared explanatory models in terms of their ability to explain charges per day, relative to costs per day. They thought that this type of comparison would be informative, given the wide range of CCRs that were calculated from the cost reports. Finally, in all of the analyses, they adjusted the SNF

labor share of the cost and charge variables by the area wage index that CMS uses in its hospital and SNF payment systems.

## **G. Explanatory Variables**

For the analysis of SNF costs and charges, they obtained explanatory variable data from 4 sources: the DataPRO stay files, the MDS, additional data from the SNF claims, and prior or qualifying hospital claims. The potential explanatory variables included demographics, primary diagnoses, service indicators, comorbid conditions (i.e., secondary diagnoses), functional status indicators and facility characteristics. Some variables, such as clinical diagnoses, were obtained from multiple sources. In order to get a better understanding of the general characteristics of SNFs and SNF residents, they provide some informative summary statistics at various levels below:

### ***Stay characteristics***

SNF residents have an average age of 80.0 years, and women incur nearly two-thirds of stays (65.9 percent). The average stay is 24.3 days. Most SNF stays are not prior nursing home residents (84.1 percent). Further, most SNF stays do not result in resident nursing home placement (84.4 percent).

### ***Clinical characteristics***

SNF residents are medically complex. The average length of stay in the qualifying hospital is 9.15 days, as compared to the Medicare average for all acute hospital stays in 2001 of 6.1 days (MedPAC June 2003). Over a quarter of SNF stays, 26.6 percent, have congestive heart failure, 23.7 percent are diabetic, and 23.6 percent have Chronic Obstructive Pulmonary Disease (COPD). Converting diagnosis codes into Major Diagnostic Categories (MDCs), at least 5 categories are present in 58.3 percent of stays. Approximately one-half of SNF admissions have cognitive impairment (49.3 percent) and more than one-third have at least moderate impairment (35.7 percent). Functionally, 54.5 percent of SNF residents are totally dependent in at least 1 of 10 ADLs at some point during their stay, and 28.7 percent are totally dependent in at least 3.

Along with functional status, the MDS provides information on treatments and procedures received by residents and on clinical conditions related to their current health status. Finally, input from clinicians led to the creation of variables to capture some potential high cost residents. These include variables that identify residents with solid organ transplants and stroke patients with poor function.

### ***Characteristics of the qualifying hospital stay***

Medicare claims data from the qualifying hospital stay provided diagnosis codes, as well as indicators and charges for a number of services. To the extent that these services must be continued in the SNF and that some services, such as intensive care unit (ICU) stays, represent sicker patients, information from the qualifying hospital may be highly predictive of costs for the SNF stay. For example, 21.1 percent of SNF stays involve stays in the ICU, and 14.6 percent of SNF residents received intravenous drug therapy during their qualifying hospital stay.

### *Facility characteristics*

The DataPRO SNF stay file also provided data on some of the characteristics of SNFs for each stay. In our sample, 78.7 percent of stays took place in facilities located in a Metropolitan Statistical Area (MSA), while 23.4 percent were in hospital-based facilities and 60.2 percent were in facilities that were part of a multi-facility system. Three-fifths of stays (61.3 percent) occurred in for-profit SNFs, 34.4 percent in non-profit SNFs, and less than two percent in government-owned facilities.