Development of Updated Models of Non-Therapy Ancillary Costs

A memo by staff from the Urban Institute for the Medicare Payment Advisory Commission

Doug Wissoker
A. Bowen Garrett
Urban Institute

•

MedPAC
601 New Jersey Avenue, NW
Suite 9000
Washington, DC 20001
(202) 220-3700
Fax: (202) 220-3759
www.medpac.gov

•

October 2010 • No. 10–8

THE URBAN INSTITUTE 2100 M STREET, N.W. / WASHINGTON D.C. 20037

To: Carol Carter and Mark Miller (MedPAC)

From: Doug Wissoker and A. Bowen Garrett (Urban Institute)

Date: June 30, 2010

Subj: Development of Updated Models of Non-Therapy Ancillary Costs (Final Memo, Order #E4058949)

Introduction

In this memo, we report on the development of updated models of non-therapy ancillary (NTA) costs for skilled nursing facilities (SNFs). We build on a model that we constructed for MedPAC in 2008. That model was estimated using a national sample of 2003 SNF stays. In that model, predicted per day NTA costs are a function of indicators of SNF care, patient functionality, hospital diagnoses, Resource Utilization Groups, and duration of stay. The model and variables are described in detail in Garrett and Wissoker (2008).

The primary purpose of this work is to develop a model that more closely meets CMS's published criteria for an administratively feasible prospective payment system.[1] CMS published the criteria as part of its proposed rule in 2009:

> We believe an administratively feasible approach to prospective payments for NTA costs would incorporate the following criteria:
> - Uses information from available administrative data (data currently required on claims or on the MDS);
> - Is case-mix adjusted, using predictor variables that represent clinically meaningful correlates of NTA services and that do not promote undesirable incentives for providers;
> - Is developed from recent data in the National Claims History, in order to assure it reflects current care patterns and practices;
> - Results in an add-on NTA index to the refined RUG case-mix groups that we are proposing based on the STRIVE project;
> - Uses a minimal number of payment groups, or levels, to limit the complexity of the SNF PPS as a whole; and
> - Ideally, uses payment groups that are clinically intuitive and readily understandable.

The original model includes elements that are at odds with two of these criteria. First, the use of hospital diagnoses in the 2008 model clearly contradicts the first criterion that payment be based on information from available administrative data. In addition, the model includes an indicator of the number of regular assessments performed for the entire stay as a measure of length of stay. This information would not always be known at the time when SNFs submit claims for payment.

[1] See Centers for Medicare & Medicaid Services, 2009, p. 22239.

Finally, the fifth criterion indicates that payment should be based on a minimal number of payment groups. The 2008 model is based on nearly 70 indicators, while a minimal model would include far fewer variables or groups.

To update the model, we use data files for 2007 SNF stays that were produced by CMS staff for its own modeling efforts. These files contain estimated NTA costs and data on the key predictors for each claim. To enable us to model average costs per stay, we aggregated the claims data to stays and used this file for our analyses.

Using this file, we first replicated our earlier model to assess whether the model retained its relatively good predictive power when estimated on the 2007 data. We found a small reduction in predictive power: The 2007 model accounts for 21 percent of the variation in the per diem NTA cost of stays as compared with 23 percent accounted for by the 2003 model. Furthermore, estimation of the same model with data measured by claim rather than by stay led to similar coefficients, but a dramatic apparent reduction in the reported predictive ability of the model from 21 to 10 percent. In CMS's analysis using a claims-based approach, they also found that the explanatory power of the model was much lower than we had previous found. However, further analysis showed that by aggregating the claim-level model predictions to the stay-level to produce a consistent measure of model fit, one can maintain the higher predictive power of the model.

Next, we adjusted the model specification to replace the variables that would be difficult for claims contractors to operationalize and for SNFs to estimate their revenues. Hospital diagnoses were replaced with diagnoses from SNFs. The last assessment on the stay was replaced with a comparable measure that can be interpreted as having a claim with a given assessment. These changes led to a minor reduction in the model's predictive power.

We then tested the extent to which expansion of the regressions can substantially increase the model's predictive power. We added roughly 200 predictors to the updated model, including indicators for the Weighted Index Model (WIM) and the diagnoses and procedures reported on the MDS, and 84 diagnosis-based risk adjustment groups (RxHCC) reported on claims data from the previous year. This substantial increase in the number of predictors led to an increase in the predictive power of the model from 20.9 to 23.2 – a relatively small increase given the large number of predictors included. Despite this, the analysis did suggest some additional variables that we later used in our revised model.

Finally, we developed two smaller models for prediction with somewhat less predictive power than the larger model. Each model is based on 20 or fewer variables. The first model was developed by estimating a regression that included a broad set of predictors and then excluding variables that occur infrequently in the data, have fairly small effects, or have only moderate statistical significance. The resulting model, which is based on 20 variables, can predict 20.6 percent of the variation in costs. A second model with categories developed using the Classification and Regression Tree (CART) software, predicts 16.2 percent of the variation in costs. If the categories are defined by stay, the model continues to predict 15.6 percent of the variation in costs. Each model contains indicators of IV and oxygen (with conditions), as well as indicators of diagnoses and an indicator of a new entry or return entry into a SNF.

The remainder of this memo provides additional details regarding the analysis described above. First, we provide an overview of the data and methods used for this work, with more detailed explanations of features that are specific to this analysis. This is followed by a presentation of the specific findings.

Data and Methods

The data and analysis for this study are parallel to those described in the MedPAC June 2008 Report to Congress and Garrett and Wissoker (2008).

The goal of the analysis is development of a predictive model of NTA per diem costs that can be used in payment of skilled nursing facilities. Two types of models are used in this work: first, a linear regression equation relating costs to patient conditions and stay characteristics; and second, a Classification and Regression Tree model in which patients are divided into groups using these same conditions. In each type of model, data on the per-day costs of a stay and patient conditions are used to estimate the relationship between costs and patient conditions and stay characteristics. The predictions of each model can be used to create a set of payment weights that when applied against the payment base rate, would raise or lower payments.

The data for this study come from CMS internal files for 2007 SNF stays. These files, which provide information for a random half of all Medicare beneficiaries, were created by CMS staff for their own work modeling NTA costs. The unit of observation is the individual claim.

CMS Construction of a Claim-level File. The starting point for the CMS file is the set of Medicare claims for 2007 SNF stays and qualifying hospital stays. Medicare claims are the primary source of data on periods of service, types of procedures furnished, patient diagnoses, and the institution's charges for services. These data were submitted by Medicare-certified providers to Medicare intermediaries for reimbursement of Medicare-covered services.

For each claim, CMS attaches information from as many Minimum Data Set (MDS) assessment records as cover the dates of the claim. The MDS assessments are the source of information on a patient's cognitive and functional status, use of specific services (such as ventilation, intravenous medication, and oxygen), and assignment to the RUG-53 category. In addition, CMS uses them as a source of information on diagnoses and procedures. The MDS is administered to patients on a specified schedule approximately 5, 14, 30, 60, and 90 days from the start of the Medicare-covered SNF stay. For a given claim, no assessment, one assessment, or multiple assessments might cover some period of the claim.

CMS staff devised an approach to attach MDS information for the one or more assessments that cover the period of a claim. For each day covered by a claim, CMS staff determined whether the day was within 13 days prior to the date of a given MDS assessment. This ensures that procedures such as IV medication are measured using the days that correspond to the 14-day look back period for each assessment date. If one assessment overlapped the period of the claim, then the MDS variables for that assessment are attached to the claim. If multiple assessments were observed, the MDS variables were defined using a weighted average of the

variable according to the share of days from each assessment that overlapped the period covered by the claim.

For example, to get a proportion of days with IV during a claim, CMS took all of the covered days that fall within 13 days before an assessment date and calculated the share of those days matched to an MDS indicating IV use. If a claim period overlapped a single assessment date, all of the days would be assigned the value of 0 or 1, indicating no IV use or IV use. If a claim period overlapped two target dates, the proportion could be anywhere from zero to one.

Finally, CMS used the cost report data that Medicare-participating SNFs submit annually to the fiscal intermediaries to create ancillary service cost-to-charge ratios (CCRs), which they used to convert claims data on ancillary service charges to estimated costs for those services.

Constructing a Stay-level File. Our analytic goal is to model the per diem NTA cost of individual stays. This contrasts with the CMS decision to model the per diem NTA cost of individual claims. This led us to estimate models of per-day NTA costs in which both costs and predictors (e.g., SNF care, diagnoses) are measured as the day-weighted average across all the claims for each stay. These models are estimated using a data file in which there is a single record for each stay and the data are the averages of the measures for that stay. Stays with a per diem cost over $1500 – accounting for less than a tenth of a percent of stays – are excluded from our analyses.

Modeling the relationships between costs and patient condition by stay rather than by claim improves our ability to predict cost per stay and provides a corresponding measure of fit. This is desirable, since we believe that accurate prediction of per diem costs is considerably more important for stays than for individual claims. Furthermore, by averaging per diem costs over the various claims within each stay, we eliminate some of the noise that can result when measuring costs or patient condition for claims that cover very few days.

Finally, modeling with data measured by stay enabled us to take different approaches to constructing the stay-level averages for variables from claims and from the MDS. The claims variables are averaged over all claims for the stay, while the MDS data are averaged over those claims with matched MDS data. This allows us to construct our best estimate of the stay average for each measure and estimate their relationship to the average cost for the stay.

To construct a stay-level file, we first dropped claims from the CMS analytic file that have zero payments, as well as a limited number of claims with dates that overlapped. We defined stays as groups of claims separated by fewer than 60 days, restricting the total number of days in a stay to 100 days.[2] For the cost and variables from claims (e.g., diagnoses and detailed charges), we weighted the values from separate claims by the number of covered days on each claim. For example, if a stay consisted of a 10 day and 30 day claim, we would construct the weighted average cost by applying a weight of one-fourth to the average cost from the first claim and three-fourths to the average cost from the second claim.

[2] The Medicare SNF benefit only covers 100 days per episode. To allow some leeway in the measurement of days of a stay, we kept stays with 101 or 102 days, but capped the number of days in the final claim to yield a maximum of 100 days. We dropped stays with 103 or more days.

We use a separate approach to calculate the average for variables that originate on the MDS to deal with claims that have no matched MDS data. Recall that CMS matched to each claim the information from subsequent MDS records that cover the claim period. Claims that come later in a stay are much more likely to have no matched assessment.

If a claim had no MDS data, we increased the weight on the nearest claim that does have MDS data. For example, consider a stay with three claims with equal length, the first of which shows the patient received IV medication, the second shows no IV medication, and the third has no matched MDS assessment. In this case, we would calculate the stay average giving the second stay double the weight, since it is adjacent to a claim without MDS data. The weighted average thus gives an increased weight to the MDS records located adjacent to claims without a matching MDS record. The weighted average is our best estimate of the stay-level average for the MDS variable.

Variables for the Analysis. Our model and key predictors are defined in Garrett and Wissoker (2008). Below, we provide a brief overview, including a brief discussion of variables that were not included in the earlier model.

The dependent variable for all analyses is the wage-adjusted per diem non-therapy ancillary costs. The unadjusted per diem NTA cost was calculated by CMS by combining data on charges for each stay with cost-to-charge ratios (CCRs) for each facility. The charges per stay are from Medicare claims and the CCRs are from the SNF cost reports. The estimated costs are standardized for area wages using the 2008 wage index (pre-floor) and the labor share in place in 2007.

Our starting point for predictors for this modeling effort is the set of variables used in the model based on 2003 data. The most important predictors are the measures of IV and oxygen conditions defined by indications on both claims and MDS data.[3] Other key predictors include age, indicators of SNF care, diagnoses from the prior hospital stay claims, ability to perform activities of daily living, SNF mental status, 5 broad indicators of the type of RUG category into which the patient was placed, and the total number of assessments over the stay. The complete list of variables used in the prior modeling is found in Table I.[4]

To create an updated model that that meets CMS's criteria, we replaced the measures of diagnoses based on hospital claims with those based on SNF claims. In addition, we eliminated indicators of the last assessment conducted because these would be more difficult to operationalize. Instead, for each claim, we included an indicator for the most recent assessment completed (for example, the day-14 assessment). Then, to create stay-based variables, we included indicators for each assessment (say, the 14-day assessment) and calculated the share of

[3] These measures differ slightly from those used in earlier work. Here, they are defined by presence on the particular claim (rather than any claim for the stay) and on the MDS.

[4] Two of the variables from the 2008 model are not analyzed here. The indicator of a prior non-PPS nursing home stay required data on prior MDS assessments. The indicator of organ transplant required claims data on procedure codes, which we did not have for the prior hospital stay. Prior nursing home stay had a t-statistic of 18 in the previous model and could explain a portion of the reduction in r-squared.

the entire stay's days associated with it (for example, the days covered by the 14-day stay comprise half of days of the entire stay). These modifications give estimates that are easily operationalized by paying each claim based on the last assessment observed at the time the claim is to be processed.

We expanded the model to include a large number of variables from the MDS that were investigated by CMS staff. These include the diagnoses and procedures indicated in Sections I and P of the MDS, additional cognitive measures, WIM variables, as well as the nursing CMI for a given stay. We exclude a small number of variables that, if put into a payment model, could be gameable, such as tube feeding, and the measures of oxygen without conditions.

Methods for Regression Analysis. Using the sample of SNF stays and alternative sets of explanatory variables, we estimate the regression models of NTA costs per day using Poisson regression. Poisson regression, like standard regression using a logged dependent variable, produces estimates that give less emphasis to the relatively rare very costly cases, better reflecting the center of the distribution. The coefficient estimates are interpreted in the same way as the coefficients from a logged standard regression model.

The r-squared statistic, which measures the proportion of variance explained, is obtained by regression of the per-stay average costs on the model's prediction of average costs. This follows the procedure used in our earlier work.

Method for CART Analysis. We estimate the CART model, using the "rpart" module for R, with the Poisson function. The CART estimation creates sequential splits in the sample, splitting on variables that increase the model predictive capability the most. For example, in our analysis, CART split the sample first according to whether IV was received for more than 2/3 of the stay. Among those without IV for 2/3 of the stay, it next split on whether a stay had a large share of days under a first or return assessment. Among those with IV, it next split according to whether the patient had oxygen, tracheostomy, or ventilator care.

We calculate the r-squared in two ways. First, we assign stays to a CART grouping according to the cuts suggested by the model. We then calculate the r-squared statistic of a regression of per diem cost on indicators of each group. Second, to deal with the issue of having claims and stays combine information from multiple assessments, we assign a unique CART group for each claim using easy-to-apply rules (i.e., whether a condition held for more than half of the claim) and then assign a predicted cost for the claim based on that grouping. We calculate r-squared statistic by calculating the average predicted cost for each claim, aggregating to the stay, and then regressing the per stay average costs on the model's prediction of average costs. This allows us to assess the predictive value of a simple assignment by claim, which is where the assignment would take place in practice.

Findings

We first report on the replication of our earlier model using 2007 data, followed by presentation of an updated model in which we replace diagnoses and length of stay measures with variables that are more acceptable to CMS.

We then explore whether adding a substantial number of predictors can dramatically increase the model's predictive value and whether credible predictive power can be retained when the model is reduced to between 10 and 20 variables. For these later analyses, we estimate the models using the sample for the model with the largest number of variables.

Replication using 2007 Data. In Table I, we report the findings of the model replication. The first column reports the NTA model based on 2003 data. The replication based on 2007 data is reported in the second column. The coefficients show similar patterns – the signs and approximate magnitude are generally the same – however, the precise coefficients have shifted somewhat over time. Slightly fewer than half of the 2007 estimates are statistically different from the 2003 estimates. The two sets of coefficients that show the largest changes are the coefficients on the broad case-mix groups (e.g. rehabilitation only or special care) and the coefficients on the number of assessments. For example, the first assessment had a coefficient of 0.75 in the 2003 data and now has a coefficient of 0.60. The predictive power (i.e., r-squared) of the model using 2007 data is 0.207 as compared to 0.23 using the 2003 data.

We next compare estimates using unweighted and weighted data measured by claim with those measured by stay to assess the effect of the "level" of estimation. This analysis is intended to help understand the differences between the models estimated with the CMS method and our own models. CMS staff use unweighted data measured by claim for their models, while we use data measured by stay. Our hypothesis was that weighting the claims-level model would lead to coefficients similar to those in the stay-level model, as stay-level data is effectively weighted by length of claims.

We first estimate the model parameters using unweighted data measured per claim. The model coefficients are of similar sign and magnitude; however, 37 percent of the coefficients are statistically different from the stay-level coefficients. We then estimate the model parameters, weighting by share of stay days within each claim. The resulting model, which is shown in column 4, yields coefficients quite similar to those obtained using the stay-level data. Six percent of the coefficients from this model are statistically different from the stay-level estimates. For each model, the predictions yield an r-squared of roughly 10 percent. That is, the model predictions explain 10 percent of the variation in the average per day NTA cost of claims. If instead we aggregate the predictions to the stay and then regress the average per day NTA costs of stays on the predictions, we obtain an r-squared of 0.202. That is, the model predictions explain 20 percent of the variation in the average per day NTA cost of stays, comparable to the stay-based findings.

Updated Model to Replace Variables Not Easily Operationalized. We next update the model, replacing in turn the measures not easily observable by SNFs. First, we replace the diagnoses based on prior hospital stay claims with those based on SNF claims. The new coefficients are reported in Column 5 of Table I. The reduction in the model's predictive ability is relatively small (e.g., the r-squared drops by 0.002), but the coefficients do change noticeably. Switching from hospital to SNF diagnoses leads to statistically significant changes in coefficients for 43 percent of diagnoses. In addition, we test the effect of treating certain diagnoses as chronic, so that any report of the diagnoses indicates presence of the condition. This change, reported in column 6 of Table I, led to very little change in coefficients.

Using the indicator of the most recently completed assessment, the change in overall r-squared is quite small (Table II, Column 2). However, the coefficients on the assessments do change. This is expected, since now the coefficients provide the marginal cost of each assessment rather than an average for the entire stay.

For comparison, we also present the coefficients for a parallel model estimated using the claims file (Table II, Column 3). In this model, the assessment is measured as a simple indicator of the last assessment attached to each claim. The results show that the coefficients are generally similar to those estimated by stay, with the exception of the coefficient on the third assessment, which is larger when the model is estimated by stay.

Models Adding Other Potential Regressors. We investigate the potential predictive ability of model based on SNF administrative data by adding approximately 200 additional independent variables to the updated model. The updated model contains nearly 70 variables available to SNFs. We add the nursing CMI, MDS measures of diagnoses, procedures, and mental status, and Weighted Index Model variables, which had been evaluated by CMS staff. In addition, we add the 84 RxHCC group identifiers based on the previous year's claims, as well as additional claims diagnoses.

The "all variables" model contains 269 predictors and yields an r-squared 0.232. Excluding the 84 RxHCC indicators reduces the r-squared to 0.226. (See Table III, models 1 and 2). These findings suggest that although one can improve on the r-squared from our updated model, the improvement is not very large and requires a large number of predictors.

Models with Few Variables. In this section, we focus on development of models with reasonable predictive power that rely on relatively few variables. The goal is to address CMS's preference for a relatively small number of casemix groups or a predictive equation based on relatively few measures. We discuss three approaches: 1) Assigning cases into one of 10 groups based on predictions from a more complicated model; 2) Regression modeling; and 3) Classification and Regression Tree models.

A simple approach to obtaining relatively few groups is to divide stays into 10 groups based on the predicted values of a relatively complicated model. The groups are based on the following percentile ranges of the predicted costs: $<1^{st}$ percentile, $1 - 5^{th}$, $5^{th} - 10^{th}$, $10^{th} - 25^{th}$, $25^{th} - 50^{th}$, $50^{th} - 75^{th}$, $75^{th} - 90^{th}$, $90^{th} - 95^{th}$, $95^{th} - 99^{th}$, and $> 99^{th}$. The findings give a sense of what is lost if one were limited to only 10 payment categories. We report the results in the last column of Table III for several of the larger models described above.

The limited number of payment categories reduces the predictive power of the model, but not by as much as one might think. For instance, the 10-group version of our updated model obtained an r-squared 0.200 as compared with 0.208 from the full model (see Table III, model 0). For the "all variables" model in which the prediction is based on 269 variables, indicators for the 10 groups yield an r-squared of 0.225 as compared with 0.232 from the full model (see Table III, model 1).

The regression and CART models attempt to meet the CMS goal of having the payment weights rely on no more than twenty variables. For these models, we considered as potential predictors or groupers variables that either were included in our updated model or were suggested by CMS staff (see Table III, model 2).

To select variables for 20-variable regression model, we began with the model that used all 185 potential predictors. We re-estimated the model, excluding independent variables that had either only moderate significance or a small coefficient. We then re-estimated the model several times, excluding indicators of conditions that are relatively rare and those that in the smaller model had relatively small coefficients.

The model, which is reported in Table IV, has 20 independent variables and yields an r-squared of 0.206. The strongest predictors were also the cornerstone of our earlier work: IV medication and oxygen/trachestomy/ventilator reported on both claims and the MDS assessment. Also important was an indicator of the proportion of stay spent in each assessment number. From the variables proposed by CMS, the nursing case-mix index and the MDS diabetes diagnosis proved to be particularly strong predictors. Other variables in the model include: age, age squared, presence of a 4^{th} stage ulcer, lack of locomotion on the unit, MDS indicator of chemotherapy, pneumonia, and claims diagnoses for COPD, infectious and parasitic diseases, respiratory failure, renal failure, and HIV.

Experimentation with the model suggests that several variables could be eliminated without substantial loss of predictive value. For example, we can drop five variables – HIV, the interaction of IV medication and oxygen/tracheostomy/ventilator, chemo, pneumonia, and age squared – and reduce the r-squared by 0.004, from 0.206 to 0.202.

Finally, we used the CART procedure to define classification groups. The CART procedure resulted in 11 groups, which are detailed in Figure 1. The 11 groups are based on the interactions of seven variables, all of which are found to be important predictors in the 20-variable regression model. The seven variables are: IV medication, oxygen/tracheostomy-/ventilator, age, the nursing case-mix index, diabetes, COPD, and proportion of the stay with an MDS indicating first assessment or re-entry.

Applying the groups defined by CART for each stay, we obtain an r-squared of 0.162. If we instead apply groups at the claim level, we obtain an r-squared of 0.157. Whether this is sufficient depends on the relative desirability of improvements in r-squared and having few regressors.

Summary

Using 2007 data for SNF stays provided by CMS to re-estimate our model of NTA per diem costs, we obtain estimates that are qualitatively similar to those obtained using 2003 data. The findings are similar in predictive ability and in the effects of specific variables. The most notable differences are those for resource utilization groups and the length of the stay. We then revise the model to exclude variables based on the qualifying hospital stay or known only at the end of a SNF stay, and to consider some new variables that were suggested by the work of CMS staff.

To reflect CMS's preference for models that use fewer than the nearly 70 variables used in our earlier model, we developed a regression model that achieves an r-squared of 20 percent using 20 variables. Finally, we produced a range of models that result in only 10 or 11 casemix groups, which have r-squared values ranging from 0.157 to 0.225. These findings suggest that a range of options are available that would greatly improve the accuracy of Medicare payments to SNFs for NTA services and present minimal administrative burden in implementation.

References

Centers for Medicare & Medicaid Services, Department of Health and Human Services. *Medicare Program; Prospective Payment System and Consolidated Billing for Skilled Nursing Facilities for FY 2010; Minimum Data Set, Version 3.0 for Skilled Nursing Facilities and Medicaid Nursing Facilities; proposed rule. Federal Register* 74, no 90:22208 – 22316, 2009.

Garrett, Bowen and Wissoker, Douglas A, *Modeling Alternative Designs for a Revised PPS for Skilled Nursing Facilities: Final Report for the Medicare Payment Advisory Commission*. Washington, DC: MedPAC, 2008. http://www.urban.org/publications/411706.html.

Medicare Payment Advisory Commission, *Report to the Congress: Reforming the Delivery System.* Washington, DC: MedPAC, 2008.

Table I. Comparison of 2003 and 2007 Coefficients for UI Model of Per Diem NTA Costs (test statistics in italics)

	2003	2007 Stay-based estimates, with diagnoses based on hospital claims	2007 Claims-based analysis, un-weighted, with diagnoses based on hospital claims	2007 Claims-based analysis, weighted, with diagnoses based on hospital claims*	2007 Stay-based estimates, with SNF claim diagnoses	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic
Age						
Age > 50 (indicator)	-0.08139	-0.05459	-0.04579	-0.05865	-0.05167	-0.05202
	-2.68	*-3.74*	*-3.14*	*-4.04*	*-3.52*	*-3.54*
Age - 50, capped at 45 = 95 - 50	-0.00312	-0.002571	-0.001046	-0.002356	-0.001654	-0.001659
	-1.95	*-3.32*	*-1.33*	*-3.06*	*-2.13*	*-2.14*
(Age - 50) squared, capped at 45^2	-0.0001737	-0.0002065	-0.0002273	-0.0002137	-0.0002222	-0.0002216
	-6.13	*-15*	*-16.33*	*-15.66*	*-16.08*	*-16.03*
SNF Care						
IV medication (MDS) *and* claim for IV therapy or solution	0.7358	0.7517	0.6838	0.7354	0.7376	0.7378
	48.45	*60.72*	*64.2*	*65.45*	*60.84*	*60.85*
IV medication*Oxygen/tracheotomy/ventilator	-0.2947	-0.1996	-0.1088	-0.1753	-0.2158	-0.2158
	-10.85	*-7.74*	*-4.47*	*-7.25*	*-8.67*	*-8.66*
IV medication*SNF MDC for respiratory	-0.06985	-0.02987	-0.03409	-0.03581	-0.04085	-0.04073
	-3.81	*-2.37*	*-2.98*	*-3.15*	*-3.21*	*-3.2*

	2003	2007 Stay-based estimates, with diagnoses based on hospital claims	2007 Claims-based analysis, un-weighted, with diagnoses based on hospital claims	2007 Claims-based analysis, weighted, with diagnoses based on hospital claims*	2007 Stay-based estimates, with SNF claim diagnoses	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic
Oxygen (linked to conditions) or tracheotomy care or ventilator and claim for respiratory or pulmonary	0.524	0.5596	0.4488	0.5359	0.5335	0.5331
	28.61	*29.1*	*22.82*	*28.79*	*29.24*	*29.24*
MDC, respiratory	0.1178	0.1116	0.08361	0.108	0.0703	0.0672
	12.7	*19.46*	*16.05*	*19.81*	*11.08*	*10.73*
No infection	-0.09029	-0.02856	-0.02784	-0.03313	-0.03686	-0.03708
	-11.18	*-3.43*	*-3.75*	*-4.21*	*-4.5*	*-4.53*
Ulcer	0.1313	0.1361	0.09861	0.1217	0.1445	0.1448
	8.72	*14.67*	*11.95*	*14.22*	*15.09*	*15.11*
Shortness of breath	0.07767	0.08195	0.07577	0.07886	0.1057	0.1051
	7.72	*13.38*	*13.05*	*13.4*	*17.27*	*17.15*
Surgical Wounds	0.03381	0.06452	0.07837	0.06666	0.06562	0.06548
	3.94	*13.57*	*18.6*	*15.02*	*14.05*	*14.01*
Chewing problem	-0.03507	-0.04378	-0.05764	-0.0472	-0.03997	-0.0401
	-3.85	*-7.8*	*-10.44*	*-8.69*	*-7.18*	*-7.21*
Swallowing problem	0.01747	0.02265	-0.001491	0.01495	0.02871	0.02886
	1.78	*4.35*	*-0.29*	*2.96*	*5.56*	*5.59*
SNF CPS Score						
Borderline	-0.0236	0.02366	0.06755	0.03935	0.02415	0.02402
	-2.22	*3.6*	*11.08*	*6.39*	*3.67*	*3.65*
Mild Impairment	-0.05651	-0.01622	0.03574	0.005728	-0.02047	-0.02067

	2003	2007 Stay-based estimates, with diagnoses based on hospital claims	2007 Claims-based analysis, un-weighted, with diagnoses based on hospital claims	2007 Claims-based analysis, weighted, with diagnoses based on hospital claims*	2007 Stay-based estimates, with SNF claim diagnoses	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic
	-5.43	*-2.52*	*6.12*	*0.95*	*-3.2*	*-3.23*
Moderate impairment	-0.08344	-0.01364	0.05126	0.01057	-0.02598	-0.0264
	-7.57	*-2.05*	*8.38*	*1.69*	*-3.93*	*-3.99*
Moderate severe impairment	-0.1443	-0.06903	-0.00172	-0.04033	-0.08744	-0.08771
	-8.54	*-6.96*	*-0.19*	*-4.37*	*-8.82*	*-8.85*
Severe impairment	-0.07091	-0.05403	0.01298	-0.02881	-0.0773	-0.07817
	-3.51	*-4.51*	*1.19*	*-2.59*	*-6.45*	*-6.53*
Very severe impairment	-0.1663	-0.1492	-0.06308	-0.1158	-0.1726	-0.1729
	-8.28	*-10.93*	*-5.22*	*-9.24*	*-12.67*	*-12.69*
Prior non-pps nursing stay	-0.2004					
	-18.1500					
SNF ADLs						
Transfer(self)-supervision	-0.02904	-0.02528	-0.04629	-0.02534	-0.02436	-0.02429
	-1.7	*-2.55*	*-4.85*	*-2.73*	*-2.45*	*-2.45*
Transfer(self)-limited assistance	-0.06133	-0.07689	-0.1225	-0.0865	-0.07714	-0.07696
	-3.66	*-8.19*	*-13.87*	*-9.82*	*-8.21*	*-8.19*
Transfer(self)-extensive assistance	-0.04002	-0.06508	-0.1006	-0.0691	-0.05873	-0.05856
	-2.15	*-6.29*	*-10.36*	*-7.09*	*-5.72*	*-5.71*
Transfer(self)-total dependence	-0.007522	-0.06247	-0.08505	-0.06644	-0.04539	-0.04489
	-0.35	*-4.68*	*-7.25*	*-5.35*	*-3.46*	*-3.42*
Transfer(self)-dir not occur during entire 7 days	-0.01813	-0.09562	-0.1289	-0.1061	-0.07488	-0.07449
	-0.56	*-4.2*	*-6.41*	*-4.97*	*-3.32*	*-3.31*

	2003	2007 Stay-based estimates, with diagnoses based on hospital claims	2007 Claims-based analysis, un-weighted, with diagnoses based on hospital claims	2007 Claims-based analysis, weighted, with diagnoses based on hospital claims*	2007 Stay-based estimates, with SNF claim diagnoses	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic
Locomotion on unit(self)-supervision	-0.003627	0.00562	-0.005054	0.003489	-0.003325	-0.003422
	-0.23	*0.59*	*-0.57*	*0.39*	*-0.35*	*-0.36*
Locomotion on unit(self)-limited assistance	0.03486	0.03268	0.02258	0.02994	0.02308	0.02304
	2.37	*3.78*	*2.77*	*3.67*	*2.7*	*2.69*
Locomotion on unit(self)-extensive assistance	0.08109	0.05243	0.052	0.05276	0.04861	0.04863
	4.53	*5.25*	*5.61*	*5.6*	*4.93*	*4.92*
Locomotion on unit(self)-total dependence	0.1113	0.1336	0.1243	0.1337	0.1317	0.1318
	6.61	*10.18*	*11.71*	*10.88*	*10.3*	*10.3*
Locomotion on unit(self)-dir not occur during entire 7 days	0.3063	0.3006	0.2785	0.2973	0.2822	0.2823
	14.27	*16.54*	*18.75*	*17.28*	*16.18*	*16.2*
Eating(self)-supervision	0.03147	-0.0003856	-0.001551	-0.002235	0.003415	0.003476
	2.79	*-0.04*	*-0.19*	*-0.27*	*0.39*	*0.39*
Eating(self)-limited assistance	0.0871	0.01573	-0.01064	0.008488	0.01571	0.01588
	6.22	*1.53*	*-1.17*	*0.88*	*1.54*	*1.56*
Eating(self)-extensive assistance	0.1053	0.02444	-0.00366	0.01605	0.02447	0.02452
	6.26	*2.35*	*-0.41*	*1.66*	*2.39*	*2.4*
Eating(self)-total dependence	0.15	0.0843	0.03841	0.06674	0.08524	0.08529
	8.91	*7.71*	*3.98*	*6.48*	*7.91*	*7.91*
Eating(self)-dir not occur during entire 7 days	0.007386	-0.2827	-0.2607	-0.2932	-0.2698	-0.27
	0.14	*-6.59*	*-6.5*	*-7.05*	*-6.33*	*-6.34*

	2003	2007 Stay-based estimates, with diagnoses based on hospital claims	2007 Claims-based analysis, un-weighted, with diagnoses based on hospital claims	2007 Claims-based analysis, weighted, with diagnoses based on hospital claims*	2007 Stay-based estimates, with SNF claim diagnoses	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic
Diagnoses						
Sepsis	-0.005546	-0.03929	-0.05063	-0.03901	-0.05161	-0.0513
	-0.38	*-4.77*	*-6.54*	*-5.03*	*-3.3*	*-3.28*
Cellulitis	0.05406	0.05582	0.04793	0.05855	0.02969	0.02955
	3.45	*6.17*	*5.8*	*6.89*	*2.26*	*2.25*
Malnutrition	0.07722	0.04528	0.03216	0.04347	0.1326	0.1323
	6.03	*6.41*	*4.85*	*6.49*	*8.23*	*8.21*
Mental disorders	-0.009995	-0.05996	-0.04043	-0.05355	-0.002669	0.00263
	-1.18	*-14.42*	*-9.62*	*-13.34*	*-0.53*	*0.55*
Hip fracture	0.01813	0.04102	0.04163	0.04689	-0.003811	-0.003891
	1.36	*6.26*	*6.35*	*7.43*	*-0.41*	*-0.42*
Stroke	-0.07553	-0.05556	-0.0446	-0.04904	-0.06109	-0.06035
	-5.79	*-7.17*	*-5.96*	*-6.69*	*-7.11*	*-7.03*
Respiratory infection	0.03711	0.03169	0.03638	0.03061	0.04271	0.04424
	4.1	*6*	*7.7*	*6.21*	*5.55*	*5.74*
COPD	0.1105	0.114	0.1225	0.1154	0.1179	0.1202
	13.69	*25.15*	*28.39*	*26.54*	*17.51*	*19.21*
Dementia	-0.07758	-0.0559	-0.05888	-0.05915	-0.06041	-0.05904
	-7.41	*-10.56*	*-11.15*	*-11.65*	*-10.06*	*-10.43*
Osteoarthritis	-0.0803	-0.1026	-0.09441	-0.09972	-0.06738	-0.05779
	-7.4	*-16.07*	*-15.12*	*-16.12*	*-8.16*	*-7.56*
Osteoporosis	-0.02233	-0.02918	-0.01145	-0.02683	-0.01651	-0.01064

	2003	2007 Stay-based estimates, with diagnoses based on hospital claims	2007 Claims-based analysis, un-weighted, with diagnoses based on hospital claims	2007 Claims-based analysis, weighted, with diagnoses based on hospital claims*	2007 Stay-based estimates, with SNF claim diagnoses	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic
	-1.82	*-4.72*	*-1.79*	*-4.5*	*-2.24*	*-1.57*
Renal failure	0.09187	0.1163	0.09754	0.1106	0.1403	0.1409
	9.52	*28.09*	*24.75*	*27.84*	*20.02*	*20.1*
Respiratory failure	0.08651	0.06797	0.07188	0.06656	0.1462	0.1472
	7.74	*11.24*	*12.24*	*11.79*	*10.71*	*10.77*
Infectious and parasitic diseases	0.08284	0.1052	0.09408	0.0994	0.1886	0.1887
	9.02	*20.73*	*20.1*	*20.85*	*23.41*	*23.43*
Neoplasms	0.09078	0.07124	0.0642	0.06695	0.06554	0.06625
	8.61	*12.31*	*11.72*	*12*	*10.12*	*10.76*
Diseases of the circulatory system	0.07801	0.01839	0.02669	0.02071	0.03142	0.03085
	9.18	*4.75*	*6.9*	*5.57*	*5.81*	*5.69*
Diseases of the digestive system	0.03679	0.01241	0.009596	0.01109	0.0008806	0.0002868
	5.39	*3.68*	*2.98*	*3.45*	*0.19*	*0.06*
Diseases of the skin and subcutaneous tissue	0.06667	0.05318	0.03892	0.04547	0.08034	0.08081
	5.88	*7.86*	*6.35*	*7.22*	*7.41*	*7.44*
Diseases of the musculoske system and connective tissue	0.0348	0.01772	0.02371	0.02169	-0.002716	-0.005423
	4.04	*3.96*	*5.34*	*5.1*	*-0.4*	*-0.81*
Injury and poisoning	0.04586	0.03678	0.03857	0.03911	0.04775	0.04801
	5.79	*9.44*	*10.03*	*10.56*	*8.19*	*8.24*
Organ transplant	0.7534					

	2003	2007 Stay-based estimates, with diagnoses based on hospital claims	2007 Claims-based analysis, un-weighted, with diagnoses based on hospital claims	2007 Claims-based analysis, weighted, with diagnoses based on hospital claims*	2007 Stay-based estimates, with SNF claim diagnoses	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic
	2.5900					
Myeloproliferative	0.0361	0.1148	0.0787	0.0953	0.0883	0.0871
	3.3800	*3.1700*	*2.2900*	*2.8000*	*2.0400*	*2.0100*
HIV	0.48	0.3862	0.406	0.3775	0.2685	0.2695
	5.53	*7.7300*	*8.75*	*7.98*	*5.93*	*5.95*
Number of assessments						
One	0.7461	0.5975	0.434	0.5841	0.6108	0.6109
	48.87	*64.37*	*46.6*	*64.59*	*65.65*	*65.6*
Two	0.3405	0.2629	0.257	0.2705	0.2784	0.278
	22.91	*32.47*	*30.75*	*33.87*	*34.39*	*34.32*
Three	0.1695	0.1556	0.1999	0.1688	0.1711	0.1704
	11.85	*22.05*	*26.85*	*24.28*	*24.19*	*24.1*
Four	0.06516	0.06376	0.09169	0.06616	0.07249	0.07197
	4.3	*9.86*	*12.66*	*10.22*	*11.19*	*11.1*
Readmission		0.485	0.3571	0.4867	0.5125	0.5118
		18.89	*13.23*	*18.66*	*19.86*	*19.83*

	2003	2007 Stay-based estimates, with diagnoses based on hospital claims	2007 Claims-based analysis, un-weighted, with diagnoses based on hospital claims	2007 Claims-based analysis, weighted, with diagnoses based on hospital claims*	2007 Stay-based estimates, with SNF claim diagnoses	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic
Broad Rug Category						
Rehabilitation only	-0.3102	0.06604	0.1763	0.08242	0.08467	0.08632
	-8.49	*3.54*	*11*	*4.92*	*4.52*	*4.61*
Rehabilitation and extensive services	-0.1452	0.1438	0.3813	0.1986	0.1718	0.1732
	-3.82	*7.32*	*22.56*	*11.24*	*8.74*	*8.82*
Extensive services only	0.0317	0.2121	0.3918	0.2505	0.2342	0.2351
	0.82	*9.32*	*20.94*	*12.3*	*10.39*	*10.44*
Special care	-0.0439	0.1756	0.1687	0.1666	0.187	0.1882
	-1.13	*8.13*	*9.18*	*8.53*	*8.66*	*8.72*
Clinically complex	-0.1083	0.02959	0.004349	0.01374	0.04205	0.04282
	-2.57	*1.45*	*0.25*	*0.76*	*2.05*	*2.09*
Constant	3.9438	3.7211	3.7576	3.714	3.7047	3.702
	85.41	*150.27*	*161.6*	*158.41*	*145.7*	*145.64*
N	173,441	631,110	1,160,913	1,160,913	631,138	631,138
R-squared statistic	.23**	0.2074	0.1007	0.0977	0.2055	0.2056
R-squared statistic at stay level			0.1965	0.2024		

* Each claim is weighted by the length of claim as a share of the length of the stay.

** As reported in MedPac's June 2008 Report to Congress

Table II. Comparison of 2007 Coefficients for UI Model of Per Diem NTA Costs (test statistics in italics)

	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, last assessment	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment	2007 Claim-based estimates, weighted, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment
Age			
Age > 50 (indicator)	-0.05202	-0.05289	-0.05596
	-3.54	*-3.59*	*-3.8*
Age - 50, capped at 45 = 95 - 50	-0.001659	-0.001591	-0.001409
	-2.14	*-2.05*	*-1.82*
(Age - 50) squared, capped at 45^2	-0.0002216	-0.0002224	-0.000227
	-16.03	*-16.08*	*-16.52*
SNF Care			
IV medication (MDS) *and* claim for IV therapy or solution	0.7378	0.7418	0.7412
	60.85	*60.78*	*66.5*
IV medication & Oxygen/tracheotomy/ventilator	-0.2158	-0.2174	-0.1928
	-8.66	*-8.64*	*-8.15*
IV medication & SNF MDC for respiratory	-0.04073	-0.0376	-0.04734
	-3.2	*-2.93*	*-4.09*
Oxygen (linked to conditions) or tracheotomy care or ventilator and claim for respiratory or pulmonary	0.5331	0.5406	0.5201
	29.24	*29.49*	*29.32*

	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, last assessment	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment	2007 Claim-based estimates, weighted, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment
MDC, respiratory	0.0672	0.07003	0.06831
	10.73	*11.19*	*11.24*
No infection	-0.03708	-0.03532	-0.03748
	-4.53	*-4.29*	*-4.82*
Ulcer	0.1448	0.1393	0.1318
	15.11	*14.48*	*14.74*
Shortness of breath	0.1051	0.1021	0.1021
	17.15	*16.59*	*17.22*
Surgical Wounds	0.06548	0.06582	0.06464
	14.01	*14.03*	*14.76*
Chewing problem	-0.0401	-0.04188	-0.04079
	-7.21	*-7.5*	*-7.52*
Swallowing problem	0.02886	0.0271	0.02589
	5.59	*5.23*	*5.15*
SNF CPS Score			
Borderline	0.02402	0.0233	0.02932
	3.65	*3.53*	*4.77*
Mild Impairment	-0.02067	-0.02157	-0.009947
	-3.23	*-3.36*	*-1.66*
Moderate impairment	-0.0264	-0.02631	-0.01548
	-3.99	*-3.96*	*-2.48*
Moderate severe impairment	-0.08771	-0.08745	-0.07223
	-8.85	*-8.79*	*-7.79*

	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, last assessment	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment	2007 Claim-based estimates, weighted, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment
Severe impairment	-0.07817	-0.07568	-0.06668
	-6.53	*-6.31*	*-5.96*
Very severe impairment	-0.1729	-0.1727	-0.1545
	-12.69	*-12.65*	*-12.28*
Prior non-pps nursing stay			
SNF ADLs			
Transfer(self)-supervision	-0.02429	-0.0243	-0.02008
	-2.45	*-2.44*	*-2.16*
Transfer(self)-limited assistance	-0.07696	-0.0777	-0.06715
	-8.19	*-8.23*	*-7.6*
Transfer(self)-extensive assistance	-0.05856	-0.05992	-0.04603
	-5.71	*-5.81*	*-4.72*
Transfer(self)-total dependence	-0.04489	-0.04829	-0.03528
	-3.42	*-3.65*	*-2.85*
Transfer(self)-dir not occur during entire 7 days	-0.07449	-0.07872	-0.07006
	-3.31	*-3.48*	*-3.3*
Locomotion on unit(self)-supervision	-0.003422	-0.001172	-0.005626
	-0.36	*-0.12*	*-0.63*
Locomotion on unit(self)-limited assistance	0.02304	0.02575	0.01984
	2.69	*2.99*	*2.44*

	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, last assessment	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment	2007 Claim-based estimates, weighted, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment
Locomotion on unit(self)-extensive assistance	0.04863	0.05064	0.04648
	4.92	*5.11*	*4.96*
Locomotion on unit(self)-total dependence	0.1318	0.1347	0.1258
	10.3	*10.45*	*10.38*
Locomotion on unit(self)-dir not occur during entire 7 days	0.2823	0.2904	0.2786
	16.2	*16.55*	*16.62*
Eating(self)-supervision	0.003476	0.00227	0.003175
	0.39	*0.25*	*0.38*
Eating(self)-limited assistance	0.01588	0.01482	0.01012
	1.56	*1.44*	*1.05*
Eating(self)-extensive assistance	0.02452	0.02349	0.01906
	2.4	*2.28*	*1.98*
Eating(self)-total dependence	0.08529	0.0841	0.07233
	7.91	*7.76*	*7.09*
Eating(self)-dir not occur during entire 7 days	-0.27	-0.2638	-0.275
	-6.34	*-6.17*	*-6.6*
Hospital Diagnoses			
Sepsis	-0.0513	-0.05266	-0.05715
	-3.28	*-3.34*	*-3.86*
Cellulitis	0.02955	0.02929	0.02726
	2.25	*2.22*	*2.11*

	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, last assessment	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment	2007 Claim-based estimates, weighted, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment
Malnutrition	0.1323	0.1364	0.1353
	8.21	*8.46*	*8.59*
Mental disorders	0.00263	0.0002509	-0.003037
	0.55	*0.05*	*-0.62*
Hip fracture	-0.003891	-0.00314	0.002578
	-0.42	*-0.34*	*0.28*
Stroke	-0.06035	-0.06108	-0.06208
	-7.03	*-7.1*	*-7.52*
Respiratory infection	0.04424	0.04151	0.04122
	5.74	*5.37*	*5.63*
COPD	0.1202	0.1159	0.1161
	19.21	*18.51*	*17.94*
Dementia	-0.05904	-0.05995	-0.05853
	-10.43	*-10.58*	*-10.09*
Osteoarthritis	-0.05779	-0.05858	-0.06234
	-7.56	*-7.66*	*-7.76*
Osteoporosis	-0.01064	-0.01164	-0.009059
	-1.57	*-1.71*	*-1.25*
Renal failure	0.1409	0.1399	0.1312
	20.1	*19.79*	*19.14*
Respiratory failure	0.1472	0.1458	0.1408
	10.77	*10.62*	*10.58*

	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, last assessment	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment	2007 Claim-based estimates, weighted, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment
Infectious and parasitic diseases	0.1887	0.1891	0.1802
	23.43	*23.28*	*23.04*
Neoplasms	0.06625	0.06483	0.06065
	10.76	*10.5*	*9.45*
Diseases of the circulatory system	0.03085	0.03191	0.02929
	5.69	*5.87*	*5.6*
Diseases of the digestive system	0.0002868	0.0005906	-0.000208
	0.06	*0.13*	*-0.05*
Diseases of the skin and subcutaneous tissue	0.08081	0.0809	0.08081
	7.44	*7.43*	*7.64*
Diseases of the musculoske system and connective tissue	-0.005423	-0.00536	-0.005669
	-0.81	*-0.8*	*-0.87*
Injury and poisoning	0.04801	0.04835	0.04871
	8.24	*8.27*	*8.55*
Organ transplant			
Myeloproliferative	0.0871	0.0881	0.0869
	2.0100	*2.0300*	*2.0400*
HIV	0.2695	0.2697	0.2545
	5.95	*5.96*	*5.75*

	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, last assessment	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment	2007 Claim-based estimates, weighted, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment
Number of assessments			
One	0.6109	0.7137	0.7364
	65.6	*39.91*	*65.63*
Two	0.278	0.3527	0.3443
	34.32	*20.14*	*33.59*
Three	0.1704	0.1951	0.1336
	24.1	*11.28*	*14.07*
Four	0.07197	0.05779	0.04964
	11.1	*2.91*	*5.57*
Readmission	0.5118	0.6571	0.6288
	19.83	*28.15*	*43.14*
Other		-0.04212	0.08893
Broad Rug Category			
Rehabilitation only	0.08632	0.07485	0.08006
	4.61	*3.97*	*4.78*
Rehabilitation and extensive services	0.1732	0.1671	0.1629
	8.82	*8.44*	*9.21*

	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, last assessment	2007 Stay-based estimates, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment	2007 Claim-based estimates, weighted, with SNF claim diagnoses and setting certain diagnoses to chronic, concurrent assessment
Extensive services only	0.2351	0.2309	0.2168
	10.44	*10.19*	*10.71*
Special care	0.1882	0.1842	0.1647
	8.72	*8.49*	*8.47*
Clinically complex	0.04282	0.03707	0.03203
	2.09	*1.8*	*1.77*
Constant	3.702	3.5906	3.594
	145.64	*121.74*	*145.03*
N	631,138	631,137	630,421
R-squared statistic	0.2056	0.204	
R-squared statistic at stay level			0.2

Table III. R-Squared Statistics of Alternative Models

Model Number	Description	Number of Regressors	Stay-level R-squared	10-payment-category R-squared
0	Our current model	69	0.2085	0.2002
1	SNF measures of variables included in MedPac model; plus RxHCC indicators, all MDS section I and P diagnoses and procedures, WIM measures, additional SNF diagnoses from claims	269	0.2323	0.225
2	All variables in 1., excluding RxHCC indicators	185	0.2256	0.2173
3	Model restricted to coefficients with "large enough" t-stats and non-negligible shares of the population	66	0.2229	0.2146
4	20-variable regression model	20	0.2063	
5a.	CART model (11 Groups)	10	0.162	
5b.	CART model evaluated with discrete assignment by claim	10	0.1567	
	N=574,076			

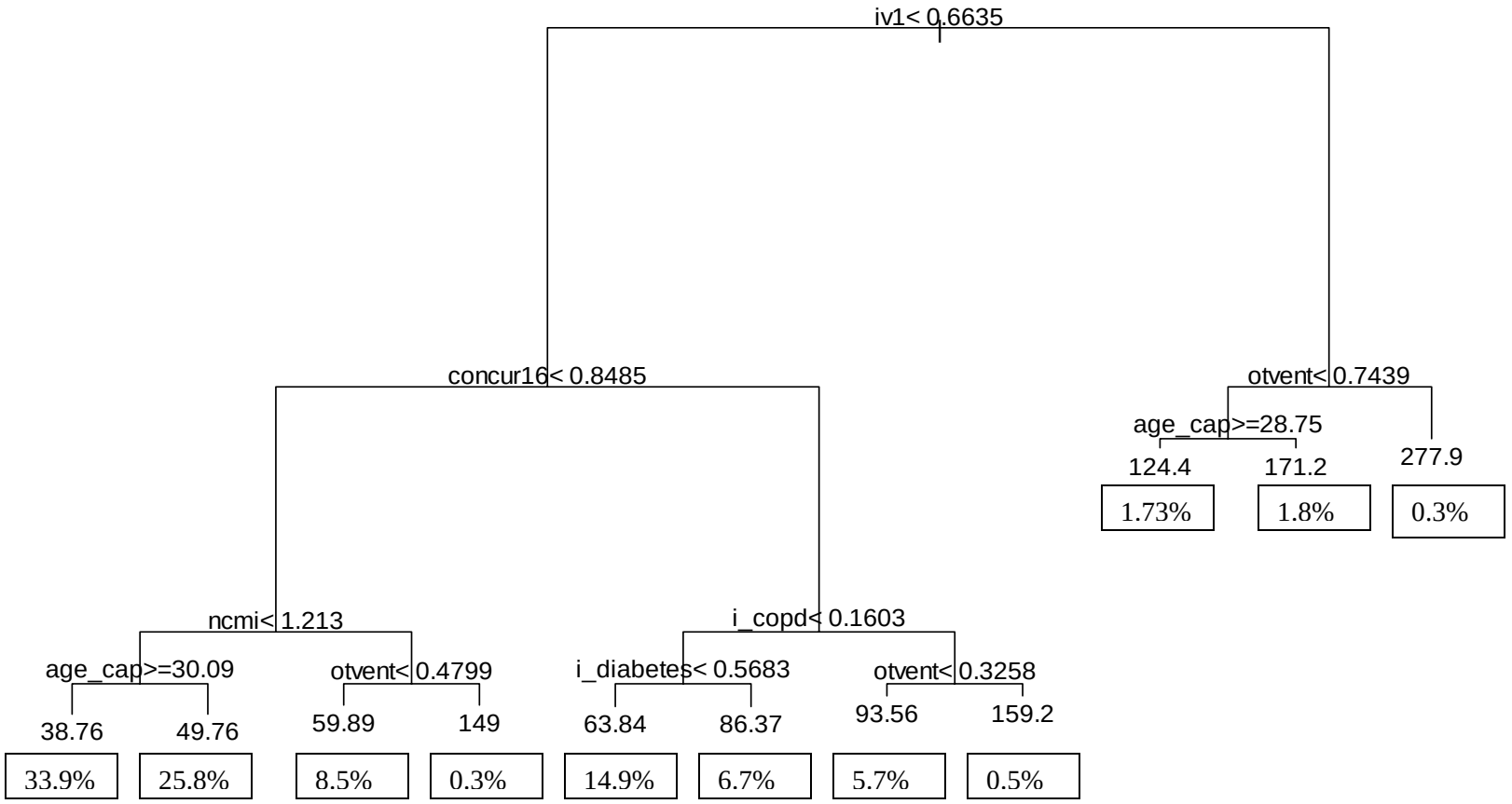
Note: These models use a common sample of stays and the results are directly comparable.

Table IV. 2007 Coefficients for Proposed Small Model of Per Diem NTA Costs (test statistics in italics) and Selected Variables for CART Model

	Small Regression Model	11-Group CART Variables
Age		
Age - 50, capped at 45 = 95 - 50	-0.0061309 *-8.70*	X
(Age - 50) squared, capped at 45^2	-0.000132 *-10.20*	
SNF Care		
IV medication (MDS) *and* claim for IV therapy or solution	0.7612 *62.89*	X
IV medication*Oxygen/tracheotomy/ventilator	-0.2529 *-9.73*	
Oxygen (linked to conditions) or tracheotomy care or ventilator and claim for respiratory or pulmonary	0.6062 *31.44*	X
Ulcer	0.1896 *20.03*	
Chemo (MDS)	0.3652 *19.73*	
SNF ADLs		
Locomotion on unit(self)-dir not occur during entire 7 days	0.2276 *15.65*	
SNF Diagnoses		
Malnutrition	0.1547 *10.49*	
COPD	0.1859 *36.64*	X (MDS)
Renal failure	0.1197 *16.18*	
Respiratory failure	0.1779 *12.49*	
Infectious and parasitic diseases	0.2030 *24.55*	
HIV	0.2959 *6.82*	

	Small Regression Model	11-Group CART Variables
Diabetes (MDS)	0.1662	X
	46.82	
Pneumonia (MDS)	0.1292	
	21.54	
Number of assessments		
One or readmission	0.6420	X
	57.13	
Two	0.3060	
	28.93	
Three	0.1509	
	14.64	
Nursing CMI	0.2933	X
	25.81	
Constant	3.4090	
	202.68	
N	574,076	574,076
R-squared statistic at stay level	0.2063	0.162
R-squared statistic defining variables by whether half of claim spent in category		0.1567
Number of variables	20	11 Groups

Figure I: CART Model of NTA Costs – Mean Cost and Population Share for each Group



KEY: Path to the left indicates that a condition is true. For example, iv1<.66 means that groups to the left have IV medication for less than 2/3 of the stay; those to the right have IV medication for more than 2/3 of the stay.

DEFINITIONS: iv1 = IV medication on MDS and claims;
concur16 = first or re-entry assessment;
otvent=oxygen/tracheostomy/vent on MDS and claims;
ncmi = nursing case mix index
	age_cap = age – 50 for 50<age<95;
	i_copd = copd diagnosis on MDS
	i_diabetes = diabetes diagnoses on MDS
		MODEL:
		N= 574,076
		R-squared = 0.162