Revision of the Case-Mix Weights for the Home Health Prospective Payment System

Final Report

November 1, 2011

Prepared for:
Centers for Medicare & Medicaid Services
7500 Security Boulevard
Baltimore, MD 21244-1850

Prepared by:
Michael Plotzke, PhD
Alan White, PhD
Henry Goldberg

Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 02138-1168

Revision of the Case-Mix Weights for the Home Health Prospective Payment System – Final Report

Table of Contents

1 Executive Summary

The purpose of the revision of the case-mix weights is to construct an improved set of payment weights using data and methods different from those used in the 2008 refinements that created the current payment weights.[1] The primary purpose of this revision is to improve the relationship between an episode's payment weight and its cost as represented by the episode's wage weighted minutes of care. The revision updates the payment regression to use episodes from 2007 instead of 2005. Also, hypertension codes 401.1 and 401.9 are removed from the four equation model and no longer impact an episode's clinical score. In addition, the revision addresses therapy incentives and comments that CMS received that therapy cases are overvalued relative to non-therapy cases. The payment weights are based on wage-weighted time spent on home health visits in our sample. The wages come from estimates of the national hourly wage for six disciplines of home health care professionals (skilled nursing, physical therapist, occupational therapist, speech language therapist, medical social services, and home health aides) from the Bureau of Labor Statistics Occupational Employment Survey.

The largest differences between the current weights and the revised weights are:

(1) The revised payment weights are derived from a payment regression equation that uses episodes from 2007 instead of 2005, the data year used for the current weights.

(2) The revised weights are created after revising the hypertension diagnosis group to exclude the ICD-9 hypertension codes 401.1 (Benign essential hypertension) and 401.9 (Unspecified essential hypertension).

(3) The revised payment regression does not control for the therapy visit ranges for an episode.

(4) Weights for episodes with a low number of therapy visits are adjusted upwards and weights for episodes with a high number of therapy visits are adjusted downwards.

The process of creating the payment weights is complex and involves a number of steps. This section provides a brief summary of the process. The remainder of the report provides a more detailed explanation of the methodology we used to create the revised case-mix weights. We also include a number of analyses designed to study how the revised set of weights compares to the current weights from several perspectives.

Four Equation Model

The first step in the revision of the case-mix weights involves estimating the four-equation model. The four-equation model is a linear regression explaining an episode's wage weighted minutes of care in the home as measured in dollars (the dependent variable) as a function of the episode's therapy visits, clinical variable indicators (e.g., pressure ulcer stage), and functional indicators (e.g., limitation in bathing). After estimating the model, we use sums of the model coefficients to impute each episode's total clinical and total functional score. We transform these scores into severity levels (low, medium, and high) using score intervals previously created during the 2008 refinements.

[1] Coleman, KA, Wu, N, Goldberg, HB, Deitz, D, and White, A. "Refinement of Medicare's Home Health Prospective Payment System: Final Report." Abt Associates, Cambridge, MA. April 2008. https://www.cms.gov/Reports/downloads/Coleman_Final_April_2008.pdf <accessed May 19, 2011>

Payment Regression

After estimating the four equation model, we assign each episode from our 2007 sample into a clinical and functional severity level (low, medium or high). At that point, we can estimate the payment regression model. The payment regression model is a linear regression model where an episode's wage weighted minutes of care (dependent variable) are regressed on indicators for the episode's step, clinical level, and functional level. Coefficients from this regression are then used to create weights for the 153 payment groups that make up the Home Health Resource Groups. The model has five "steps", which are represented by indicator variables for the following type of episodes:

- Step 1: first and second episodes, 0-13 therapy visits
- Step 2.1: first and second episodes, 14-19 therapy visits
- Step 2.2: third episodes and beyond, 14-19 therapy visits
- Step 3: third episodes and beyond, 0-13 therapy visits
- Step 4: all episodes with 20+ therapy visits

To respond to comments that CMS received regarding the large role that therapy plays in determining the case mix weights, we have changed the payment regression model to reflect a different methodology for measuring the additional resource use associated with therapy visits. We drop indicators for the number of therapy visits (which are more specific than the step variables) and rely more heavily on clinical and functional level variables. After estimating this model, we predict each episode's wage weighted minutes of care using the model's coefficients. We convert the predicted value of each episode's wage weighted minutes of care into a "raw" payment weight by dividing the predicted value by the average of the dependent variable (i.e., the average wage weighted minutes of care across all episodes used in the payment regression).

Because we did not include variables for therapy visits (beyond the step variables) in the payment regression model at this stage of the process, all episodes are assigned the same raw weight within a step (holding the clinical and functional level constant).

To achieve a gradient in the weights so that payment increases as the number of therapy visits increases, we interpolate to define weights for payment groups that represent intermediate groups of therapy visits between the steps. These are the same intermediate groups in use in the current set of weights (e.g., 7 to 9 therapy visits). We then take the interpolated weights and normalize them so that when the interpolated weights are applied to a sample of episodes from 2007 with normal episodes, Partial Episode Payments (PEPs), and Significant Change in Condition episodes (SCICs) the average equals one. Finally, we make the weights budget neutral to 2009 when using the current weights.

The remainder of the report describes this methodology in additional detail and provides analyses comparing the revised weights to the current weights generated during the 2008 refinements.

2 Methods

2.1 The Four Equation Model – Data

The first step of the revision of the case-mix weights is to estimate the four-equation model. To estimate the model we use a 20% sample of normal home health episodes from 2005. To create a sample of normal episodes, we drop episodes classified as Partial Episode Payments (PEPs), Significant Change in Condition (SCICs), Low Utilization Payment Amounts (LUPAs), Outliers, and episodes with payment amounts that were missing or were zero.[2] The 20% sample is generated by using the 0^{th} and 1^{st} decile of home health claims. Each decile of data selects ten percent of the home health using beneficiaries at random and collects all of their claims in a "decile" file. This method ensures that each decile is a sample representative of Medicare home health users nationally. Data from 2005 is selected by those with a calculated end date in 2005.[3] In 2005, there were 789,877 episodes in this 20% sample of normal episodes.

2.2 The Four Equation Model – Methods

The four-equation model is designed to identify clinical variable indicators (e.g., pressure ulcer stage), functional indicators (e.g., limitation in bathing), and service use variables that are correlated with an episode's wage weighted resource minutes. The wage-weighted resource minutes serves as the dependent variable of the four-equation model. The wages come from estimates of the national hourly wage for six disciplines of home health care professionals (skilled nursing, physical therapist, occupational therapist, speech language therapist, medical social services, and home health aide) from the Bureau of Labor Statistics Occupational Employment Survey. The minutes come from the visit times reported in the claims sample. The four-equation model is a linear regression model. After estimating the model, we use sums of the model coefficients to impute each episode's total clinical and functional score. We transform these scores into severity levels (low, medium, and high) using score intervals previously created during the 2008 refinements.

The report "Refinement of Medicare's Home Health Prospective Payment System: Final Report" provides additional detail on how the four equation model was originally constructed.[4] For the revision, we use the same variables used in the original model and use the same clinical/functional score intervals from the original model to construct an episode's clinical and functional level.[5] The score intervals were created so that in general approximately a third of episodes within each "step" were assigned to the low, medium, and high clinical and functional levels. The model has five "steps", which correspond to the following types of episodes:

- Step 1: first and second episodes, 0-13 therapy visits

---

2 Some of these omitted episodes are reinstated later in the analysis for purposes of normalizing the payment weights and ensuring budget neutrality.

3 The end date is the last date of the episode (or thru-date) if the episode was transferred to another provider (a Partial Episode Payment), otherwise the end date occurs 59 days following the first day of the episode.

4 Coleman, KA, Wu, N, Goldberg, HB, Deitz, D, and White, A. "Refinement of Medicare's Home Health Prospective Payment System: Final Report." Abt Associates, Cambridge, MA. April 2008. https://www.cms.gov/Reports/downloads/Coleman_Final_April_2008.pdf <accessed May 19, 2011>

5 The variables in both of these four equation models correspond to those shown in Table 2A of Final Rule CMS-1541-FC.

- Step 2.1: first and second episodes, 14-19 therapy visits.
- Step 2.2: third episodes and beyond, 14-19 therapy visits
- Step 3: third episodes and beyond, 0-13 therapy visits
- Step 4: episodes with 20+ therapy visits

It was not always possible to assign a third of all episodes into each level for each step. In certain cases, episodes clumped around a small number of clinical and functional scores. For example, 63% of episodes from 2005 in Step 2.2 had the same functional score. Therefore, these episodes and only these episodes were assigned to the medium functional level.

The only difference between how the 2008 refinements four equation model and the revised four equation model is estimated is that in the revised model we drop the following ICD-9 codes from the hypertension diagnosis group[6]:

- 401.1 (Benign essential hypertension)
- 401.9 (Unspecified essential hypertension)

The four-equation model was re-run without flagging episodes with those codes as members of the hypertension diagnosis group. Specifically, dropping those codes means fewer episodes are classified as having "Primary or Other Diagnosis = Heart Disease OR Hypertension." The rate at which home health agencies have coded hypertension has increased dramatically in the last few years while there has been little change in the wage weighted minutes of care for those episodes. The revised four-equation model does not assign points for these codes and therefore the model should produce a different relationship between an episode's clinical/functional score and the episode's wage weighted minutes of care during more recent years when those hypertension codes are used more frequently. Analysis of home health episodes from 2008 indicated that, after accounting for all the other conditions and therapy effects represented in our models, hypertension did not explain any marginal increase in resources. This is a different result compared to looking at home health episodes from 2005 and is likely due to the massive increase in flagging of hypertension by 2008. The definitions of all other flags used to identify conditions in the four-equation model remained unchanged.

In the revised four-equation model, certain coefficients (besides those related to hypertension) differ slightly from those estimated from the four-equation model used in the 2008 refinements. That is, some clinical and functional variables may be associated with more points, fewer points, or the same number of points after the 401.1 and 401.9 codes are dropped compared to when they are included in the model.

We show the results of the four-equation model in Table 1. Results in Table 1 are calculated by dividing the coefficients generated from the four equation model by 10 and rounding to the nearest

[6] The ICD-9 codes that make up the hypertension diagnosis group can be found in Table 2B of "Home Health Prospective Payment System Refinement and Rate Update for Calendar Year 2008; Final Rule," 72 Federal Register 167 (29 August 2007), pp. 49792–49793.

integer. Table 1 organizes the results by grouping variables across different legs of the model.[7] Table 1 only shows points for variables that contribute to an episode's clinical or functional scores.

Appendix Table 1 shows results from all variables in the model and also includes the results from the four-equation model used in the 2008 refinements. Note, these are the same results as shown in Table 1, however Appendix Table 1 also shows results from therapy visits variables that were omitted from Table 1. In that table, Model 1 contains the four-equation model results from the 2008 refinements. Model 2 estimates the model using a 20% sample of normal episodes from 2005 where the 401.1 and 401.9 codes are dropped. The results in Appendix Table 1 show some differences between the model regression results. Looking at Appendix Table 1 and the variables used to assign an episode's clinical and functional level, there were 33 variables that differed by a point between the two four-equation models. Thirteen variables had an additional point and 20 had one less point. These differences between the two models indicate that some episodes could change clinical and functional levels depending on which model is used.

Table 2 explores how episodes change clinical levels when switching between the two different four-equation models. Although it is not shown in this table, a small number of episodes also change functional groups. In the 2005 sample used for the 2008 refinements, 344 out of 789,877 episodes changed functional groups when assigned a functional score from the original four-equation model versus the revised model. For simplicity, these episodes are not included in Table 2. Each row in Table 2 indicates a payment group for the episodes that include the 401.1 and 401.9 variables. Payment groups are determined by the episode's clinical level, functional level, number of therapy visits, and whether the episodes are "early" or "later."

The column "Total Episodes (w/401.1 and 401.9)" shows the number of episodes in each payment group when using the four-equation model from the 2008 refinements. Then, the low clinical score, medium clinical score, and high clinical score columns show the number of episodes that are in each category when the 401.1 and 401.9 codes are dropped (i.e., the revised four equation model).

For example, the first row of results in Table 2 indicates that when using the four equation model from the 2008 refinements to assign clinical and functional levels using the 20% sample of episodes from 2005, there were 40,234 episodes in the lowest payment group (early episodes, 0-5 visits, low clinical level, and low functional level). Of those 40,234 episodes, 39,252 (97.6%) remained in the low clinical score category when applying the revised four equation model results. Additionally, 982 episodes (2.4%) moved to the medium clinical score category and no episodes moved to the high clinical score category. Depending on the payment group, there can be a variety of different outcomes in terms of how many episodes switch clinical levels. In some rows of Table 2, roughly 20-30% of episodes transition to other payment groups/clinical levels. In other rows, nearly all episodes remain in the same payment group/clinical level when using the revised four equation model. The column "total switchers" indicates how many episodes switched payment groups/clinical level when

[7] An episode's leg refers to whether it is an early (1st or 2nd is a sequence of episodes) or a later (3rd or higher in a sequence of episodes) episode and whether the episode had a low level of therapy (0–13 visits) or a high level of therapy (14+ visits). The different clinical and functional variables in the four equation model were interacted with an episode's leg in order to estimate the relationship between these variables and wage weighted minutes of care across legs. The interaction method is equivalent to estimating leg-specific coefficients for the clinical and functional variables.

using the revised four equation model. In 2005, 105,817 out of 789,533 (13.4%) episodes changed payment groups/clinical levels when switching between the two differing four equation models.

Figures 1 through 5 show the distribution of episodes in 2005 (using the 20% sample of normal episodes) based on their clinical and functional levels and which "step" the episode is in. Each figure includes the distribution when the 401.1 and 401.9 hypertension codes are included and also the distribution when those codes are not included. The exclusion of the hypertension codes has almost no impact on the distribution of functional score levels. The exclusion has a slightly more pronounced impact on the distribution of clinical score levels. In each step, for the distribution after removing the hypertension codes, there are more episodes in the low clinical level and fewer episodes in the medium and high clinical level compared to the distribution with the hypertension codes. However, even with the changes, the distribution of clinical levels is similar regardless of whether the hypertension codes are included. Based on these results, we recommended that CMS continue using the old score intervals (those generated during the 2008 refinements) for constructing the clinical and functional levels.

2.3 The Payment Regression

The four-equation model is used to assign episodes to a particular clinical and functional level. The payment regression from the 2008 refinements then uses that information, along with information on the episode's number of therapy visits and whether it is early (1st or 2nd is a sequence of episodes) or later (3rd or higher in a sequence of episodes), to estimate a linear regression where the dependent variable is again an episode's wage weighted minutes of care in the home. Coefficients from this regression are then used to create weights for the 153 payment groups that make up the Home Health Resource Groups.

Our first approach to the payment regression was to use our 2008 sample to estimate the exact same regression model used in the 2008 refinements, which were based on a 2005 sample. Results from this model are presented in Appendix Table 2.[8] Comparing the results in that table shows that when the regression is estimated using data from 2008, the coefficients for many of the clinical and functional variables are lower than when the regression is estimated using data from 2005. This finding, along with comments that CMS received that the model relied too heavily on therapy visits, led us to construct a revised payment regression.

Table 3 shows the estimates from the revised payment regression. The revised payment regression model differs from the current payment regression model because it does not include any indicators for the number of therapy visits during an episode. Again, this change addresses previous comments that the method of weight construction for the 2008 refinements relied too heavily on variables associated with an episode's therapy visits. By removing those variables, the payment regression model is now more dependent on an episode's clinical and functional level.

We estimated the payment regression on 20% samples of data from 2005, 2007 and 2008. Again, this sample only includes normal episodes and does not include outliers, PEPs, LUPAs, or SCICs.

Table 3 shows results of the payment regression using six different models:

[8] Note, results from Appendix Table 2 differ slightly from the corresponding payment regression in the 2008 refinements due to a different 20% sample of data compared to what was used for the 2008 refinements.

- Episodes from 2005 where **Model 1** (which **includes** the 401.1 and 401.9 codes) of the four equation model was applied to calculate an episode's clinical and functional level
- Episodes from 2005 where **Model 2** (which **does not include** the 401.1 and 401.9 codes) of the four equation model was applied to calculate an episode's clinical and functional level
- Episodes from 2007 where **Model 1** (which **includes** the 401.1 and 401.9 codes) of the four equation model was applied to calculate an episode's clinical and functional level
- Episodes from 2007 where **Model 2** (which **does not include** the 401.1 and 401.9 codes) of the four equation model was applied to calculate an episode's clinical and functional level
- Episodes from 2008 where **Model 1** (which **includes** the 401.1 and 401.9 codes) of the four equation model was applied to calculate an episode's clinical and functional level
- Episodes from 2008 where **Model 2** (which **does not include** the 401.1 and 401.9 codes)of the four equation model was applied to calculate an episode's clinical and functional level

The inclusion (or exclusion) of the hypertension codes does not cause large differences in the coefficients of the models. However, as with the model from the 2008 refinements (Appendix Table 2), the payment regression coefficients can change substantially depending on whether episodes from 2005, 2007, or 2008 are used to estimate the model. In particular, the clinical score coefficients for the medium and high levels are much lower when estimating the model using episodes from 2008 compared to 2005. Focusing on the payment regression models that exclude the hypertension codes, the differences in the clinical level coefficients for episodes in Step 2.2 (third episode or later, 14-19 therapy visits) are large. The coefficient on the medium clinical level (9-16) is $97.21 in 2005, $76.41 in 2007, and $40.19 in 2008. The coefficient on the high clinical score (17+) is $227.86 in 2005, $177.93 in 2007, and $110.78 in 2008. **These results led us to use episodes from 2007 to generate the revised payment weights**. Episodes with a higher clinical level would have a lower payment weight than they otherwise would have if we had revised using episodes from 2008 instead of episodes from 2005 or 2007. Using episodes from 2007 allowed us to use more recent data to revise the model compared to the 2008 refinements (based on 2005 data) while still maintaining a desired relationship between the clinical severity level and an episode's wage weighted resource minutes.

2.4 Methodology to Construct Payment Weights

We use the results from the fourth column of results in the payment regression from Table 3 "Episodes from 2007 (w/o 401.1 and 401.9 – Model 2)"—which uses episodes from 2007 and uses the four-equation model that excludes the hypertension codes (401.1 and 401.9)—to construct the revised payment weights. Table 4 shows the revised payment weights alongside the current payment weights for each payment group. The process for constructing these weights is described below.

(1) We use the coefficients from the payment regression model (which does not include the therapy visit range variables used in the payment regression model from the 2008 refinements) to predict each episode's wage weighted minutes of care, which is later used to help construct each episode's payment weight.[9] We then divide these predicted values by the

[9] Note, in this step and other steps in the revision of the case-mix weights we do not use the rounded results as shown in the table to perform the imputation. We instead use non-rounded results.

mean of the dependent variable (i.e., the average wage weighted minutes of care across all episodes used in the payment regression). In 2007, the mean of the dependent variable was $488.70. This division constructs the "raw" weight for each episode, which is simply the ratio of the episode's predicted wage weighted minutes of care divided by the average wage weighted resource minutes in the sample. The raw weights are constructed so that each episode is placed into one of the 153 home health resource groups and that the weights for all episodes in a particular payment group are the same.

(2) We next adjust the raw weights in step (1) to have higher values for episodes with lower therapy visits and lower values for episodes with higher therapy visits. This was done to respond to concerns that the payment weights created during the 2008 refinements are too high for the payment groups corresponding to higher therapy visits and too low for the groups corresponding to low therapy visits. That relationship could create an incentive to provide an excessive amount of therapy. The adjustments we make are designed to reduce that incentive.

In the current 153-group case weight system, each combination of early/late episodes, clinical level, and functional level has nine different therapy visit ranges used in the payment regression model to control for the number of therapy visits associated with an episode:

a. Therapy visit range 1: 0 – 5 Visits
b. Therapy visit range 2: 6 Visits
c. Therapy visit range 3: 7 - 9 Visits
d. Therapy visit range 4: 10 Visits
e. Therapy visit range 5: 11 - 13 Visits
f. Therapy visit range 6: 14 - 15 Visits
g. Therapy visit range 7: 16 - 17 Visits
h. Therapy visit range 8: 18 - 19 Visits
i. Therapy visit range 9: 20+ Visits

For the revised payment weights, we multiply the raw weights associated with the first therapy visit range by 1.0375. We multiply the raw weights from therapy visit range 6 to therapy visit range 8 by 0.975. Finally, we multiply the raw weights from therapy visit range 9 by 0.950. No direct change is made to weights associated with therapy visit range 3 to therapy visit range 5. We also tested alternative sets of adjustment factors and found these best produced a range of weights that achieved the goal of increasing the weights of the episodes with no or low numbers of therapy visits and decreasing the weights of the episodes with high numbers of therapy visits, while avoiding excessively large differences from the current weights.[10]

After applying this adjustment, the weights for a particular combination of early/late episodes, clinical level, and functional level resemble Figure 6.

[10] We tested two other variations on these adjustments (1) Multiply the raw weights associated with therapy visit range 1 by 1.15. Multiply the raw weights from therapy visit range 6–8 by 0.90. Multiply the raw weights from therapy visit range 9 by 0.80. (2) Multiply the raw weights associated with therapy visit range 1 by 1.075. Multiply the raw weights from therapy visit range 6–8 by 0.950. Multiply the raw weights from therapy visit range 9 by 0.900.

Because the payment regression has no indicators for therapy visits, the weights associated with all the therapy visit ranges (i.e., 1–9) in a given step are the same. This produces a stair step pattern as seen in Figure 6. In order to smooth this pattern and produce weights that resemble the current 153 group payment model, we interpolate the weights. That is, we "connect the dots" of the above graph in two places

- Between therapy visit range 1 and therapy visit range 6
- Between therapy visit range 6 and therapy visit range 9

This involves drawing a straight line between those points. For example, this is accomplished by regressing the weights for therapy visit range 1 and therapy visit range 6 on the values of 1 and 6. The coefficient from that regression is equal to the slope of the interpolation line. The slope and the intercept are used to construct the interpolation line and determine the interpolated weights for therapy visit range 2 through therapy visit range 5. That is, after the interpolation line is drawn, we then calculate where each therapy visit range corresponds to on that newly drawn diagonal line. A similar process is used to construct the interpolation between therapy visit range 6 and therapy visit range 9. Figure 7 provides an example of the interpolation. We perform this interpolation procedure for each set of weights and each combination of early/late episodes, clinical level, and functional level.

(3) After computing the "interpolated" weights for each payment group, we then normalize the weights from all of the payment groups so the average weight equals one when the interpolated weights from step (2) are applied to a 20% sample episodes from 2007 (which excludes the hypertension codes 401.1 and 401.9 from the scoring system) that includes normal episodes, PEPs, and SCICs. We normalize by dividing each payment group's weight by the average of all weights across all episodes in that sample. We normalize the weights to set the average weight of episodes to equal one. This has no impact on the relative values of the weight. The constant we divided by to normalize the weights was 1.2831876.[11] The weights generated at this stage correspond to the adjusted weights in Table 4.

Note, Table 4 and subsequent tables list three different weighting variations. These are referred to as:

(A) **2008 Refinement Case Mix Weights:** These weights are currently used in the Home Health Prospective Payment System and were described in the 2008 report "Refinement of Medicare's Home Health Prospective Payment System: Final Report."

(B) **Revision of Case Mix Weights Using Episodes from 2005:** These weights are constructed using the methodology described in this report. These weights were formed by estimating the payment regression on a 20% sample of normal episodes from 2005.

[11] When computing the average, we compute a weighted average, assigning a value of one to each normal and SCIC episode and a value equal to the episode length divided by 60 for PEPs.

(C) **Revision of Case Mix Weights Using Episodes from 2007:** These weights are constructed using the methodology described in this report. These weights were formed by **estimating** the payment regression on a 20% sample of normal episodes from 2007. Based on the analysis in this report, these are the recommended weights.

(4) Finally, we apply an adjustment to make the normalized weights in step (3) budget neutral to 2009. After having normalized the weights to an average case-mix of 1 using the 2007 sample, we applied the normalized weights to the 20% sample of episodes from 2009. The normalized weights from the 2007 sample resulted in an average case-mix of 1.047355 in the 2009 sample. We then multiplied those weights by a budget neutrality adjustment (which is a constant) in order to ensure that the average case-mix of our revised weights is equal to the average case-mix in 2009 under our current weights. Specifically, when looking at the 20% sample of episodes from 2009 using weights from the 2008 refinements (excluding LUPAs), the average case-mix was found to be 1.343953. Since the normalized weights result in an average case-mix of 1.047355 when applied to the 2009 sample, the weights were increased by a budget neutrality factor of 1.2831876 to make sure the average case-mix between the revised and current weights are equal (1.343953/1.047355 = 1.2831876). This step produces the final revised weights as shown in Table 4.

3 Analysis of Payment Weights

The remainder of this report examines the payment weights using a variety of different methods to compare the final weights from the 2008 refinements to the final weights from the revision of the case-mix weights.

3.1 Weights and Marginal Weights

Table 5 shows the final weights from Table 4, but the weights are now organized into groups where the payment groups have the same early/late indicator, clinical level, and functional level. This table shows the progression of weights as the therapy visit range increases from 1 to 9. Additionally, it shows the marginal change in weights from one therapy visit range to the next highest therapy visit range. By design, the marginal change in weights for the revised weights share the same value for all the therapy visit ranges corresponding to between 0 and 15 visits and also all the therapy visit ranges corresponding to between 16 and 20+ visits.

Although the revision of the case mix weights using the 20% sample of episodes from 2007 is the recommended approach, tables 4 and 5 show that differences in the weights between that approach and the revision of the case mix weights using episodes from 2005 is very slight. For example, the lowest weights for the revised weights come from the payment group corresponding to later episodes, 0-5 therapy visits, with low clinical and functional levels (Final Weight (2005): 0.6808, Final Weight (2007): 0.6692). The highest weights come from later episodes, 20+ therapy visits, with high clinical and functional levels (Final Weight (2005): 3.0120, Final Weight (2007): 3.0014). At most, the revision of case mix weights using episodes from 2007 is 3.4% higher (payment group 10131) and 6.4% lower (payment group 22321) compared to the revision of case mix weights using episodes from 2005.

The revision of the case mix weights using episodes from 2007 shows more differences when compared to the 2008 refinement weights. These differences correspond to the adjustments made to increase the weights of episodes with no or low levels of therapy and decrease the weights of episodes with high levels of therapy. With the 2008 refinement weights, the lowest weight comes from the payment group corresponding to early episodes (instead of later episodes), 0-5 therapy visits, with low clinical and functional score (2008 Refinement final weight: 0.5827, Revision of case mix weights using episodes from 2007: 0.8186). Similar to the revision of case mix weights using episodes from 2005, the 2008 refinement final weights have the highest weight corresponding to later episodes, 20+ therapy visits, with high clinical and functional levels (2008 Refinement Final Weight: 3.4872, Revision of case mix weights using episodes from 2007: 3.0014). The largest differences in weights between the two systems occur in payment group 10121 (Revision of case mix weights using episodes from 2007 is 53% higher compared to the 2008 refinement weights) and payment group 40231 (Revision of case mix weights using episodes from 2007 is 14% lower compared to the 2008 refinement weights).

3.2 Figures Comparing Final Weights

Figures 8–10 plot the weights from the 2008 refinements and the revision of the case-mix weights using data from 2005 and 2007. As in Table 5, weights are grouped by those with the same early/late indicator, clinical level, and functional level. The only thing changing within a graph is the therapy visit ranges. As seen in the figures:

- The revised weights have a flatter slope compared to the weights from the 2008 refinements.
- In most cases, the revised weights have a higher weight for the first therapy visit range and a lower weight for every additional therapy visit range in comparison to the current weights. The first therapy visit range corresponds to almost 60% of all episodes samples of data from 2005 and 2008.
- The revised weights are very similar regardless of whether episodes from 2005 or 2007 are used when estimating the payment regression.

3.3 Predictive Ratios: Predicted Resource Dollars over Actual Resource Dollars

Tables 6, 7, and 8 analyze ratios of predicted resource dollars to actual resource dollars. An episode's predicted resource dollars equals its payment weight multiplied by the average resource dollars from the year the weights were constructed. That is, to construct the predicted resource dollars using the weights from the revision of case mix weights using episodes from 2007, we multiply each episode's payment weight by $488.70 (the average wage weighted resource minutes for episodes in 2007). Actual resource dollars are equal to the episode's wage weighted resource minutes.

For each table, we adjust each episode's predicted resource dollars so that the aggregate ratio of predicted resource dollars to actual resource dollars equals one. The aggregate ratio equals the sum of all episodes' predicted resource dollars over the sum of all episodes' actual resource dollars. We make the adjustment so that we can more easily compare the impact of the different weighting variations on the ratio. Before dividing by the constant for each of the predictive ratio tables, the overall predictive ratio for each of the weighting variations is either above one or below one. This is expected because the different weighting variations produce different relationships between an episode's weight and its wage weighted resource minutes. Our adjustments are only meant to normalize the predictive ratios for each variation so that the range of the predictive ratios for each variation is more evident and comparable.

3.3.1 Predictive Ratios by Deciles of Weight

Tables 6a and 6b present predictive ratios by approximate deciles of each episode's weight. The decile groups are determined by arraying the weights associated with the episodes in the sample from smallest to largest and then subdividing them into ordered groups that have no overlapping weights. This procedure does not produce evenly sized groups, so we call these "approximate" deciles. The predictive ratio is calculated by summing (within a decile) all episode weights multiplied by the average wage weighted resource minutes for the year the weights were constructed and then dividing by the sum of the actual wage weighted resource minutes for those episodes. Table 6a applies three sets of weights (the 2008 refinements weights and the revision of the case mix weights using episodes from either 2005 or 2007) to a 20% sample of normal episodes from 2005 and computes the approximate deciles and ratios based on these data. Table 6b applies the same sets of weights to a 20% sample of normal episodes from 2008 and computes the approximate deciles and ratios based on these data.

Ratios that are above one indicate that those episodes in a particular decile have predicted wage weighted minutes of care on average that exceed their actual wage weighted minutes of care. Ratios

that are below one indicate that those episodes in a particular decile have predicted wage weighted minutes of care that are less than their actual wage weighted minutes of care.

In both Table 6a and Table 6b, using the weights from the 2008 refinements, the predictive ratios are below one for the first decile, above one for the 2nd through 6th decile and then below one for the 7th – 10th decile. Applied to episodes from 2005, the ratios from the revised weights are above one for the 1st through 7th decile when using the case mix weights using episodes from 2005 and the ratios are above one for the 1st through 6th decile when using the case mix weights using episodes from 2007. The ratios from the 2007 revised weights are below one for the other deciles. That indicates the revised weights show a slightly more stable pattern with the lower weights having a ratio above one and higher weights having a ratio below one. The range of ratios for the revised weights is larger than those for the weights from the 2008 refinements. This is because of the upward adjustments for the weights for payment groups with a low number of therapy visits and the downward adjustments for the weights for payment groups with a high number of therapy visits.

3.3.2 Predictive Ratios by Number of Therapy Visits

Tables 7a and 7b present predictive ratios by average number of therapy visits. We compute the ratios in the same manner as we did in Tables 6a and 6b, except that episodes are grouped according to the number of therapy visits. Table 7a applies the same three sets of weights that were used in Tables 6a and 6b to a 20% sample of normal episodes from 2005. Table 7b applies the same three sets of weights to a 20% sample of normal episodes from 2008.

In both Table 7a and 7b the predictive ratios from the 2008 refinements do not show a steady trend throughout the range of therapy visits. For example, in Table 7a the ratios from the 2008 refinements are above one for episodes with zero and one therapy visits. The ratios are then below one for episodes with between two and five therapy visits. Then the ratios are above one for episodes with between six and eight visits. Additionally, for episodes with around 20 therapy visits, the ratio is above one. Conversely, looking at the revised weights shows that the pattern in the ratios is much more consistent. For Table 7a, for either the revision of case mix weights using episodes from 2005 or 2007, episodes with between zero and three therapy visits and between five and eight therapy visits have a ratio above one. Episodes with eight or more therapy visits have a ratio one or below. Table 7b shows a similar relationship for the revision of case mix weights using episodes from 2005 or 2007. Episodes with between zero and three therapy visits have a ratio of one or above one. Episodes with either six or seven therapy visits also have a ratio above one. Episodes with eight or more therapy visits have a ratio below one. Again, the revised weights have been adjusted upwards for payment groups with a low number of therapy visits and downwards for payment groups with a high number of therapy visits. Also, the revised weights have a larger range of possible ratios because of these adjustments.

3.3.3 Revision Weights and the Implications for Episode Payment and Cost by Number of Therapy Visits

Although the predictive ratios are below 1 for episodes with 8 or more therapy visits, this does not imply that providers lose money on these episodes. To fully understand the profitability of the episode, the actual costs of the episode (not simply the resource cost) and the payments must be

analyzed. To do this, we examined episode payments and costs using a sample of claims from 2009 and the corresponding cost reports for those claims. We took several steps in order to only use cost reports (and the corresponding claims) with valid data and which would not be considered an outlier. Specifically, we did the following:

1. We initially used a file of cost reports from 2009 which contained 9,149 cost reports representing 8,910 unique provider numbers. This included only freestanding agencies. Facility based agency cost reports were not used in this analysis.
2. Next we attempted to isolate cost reports with extreme data. We flagged the following cost reports.
    a. Cost reports that had provider numbers that were designated as "Rehabilitation Facility Based Program", "Hospital Based Program", or "Skilled Nursing Facility Based Program" in the Provider of Services file.
    b. Cost reports that contained 10 or fewer months of data or contained 14 or more months of data (using the fiscal begin date and fiscal end date on each report).
    c. Cost reports that were missing all variables used to construct "total payments"[12]
    d. Cost reports that were missing all variables used to construct "total costs"[13]
    e. Cost reports that were in the top 1st or 99th percentile of "cost per episode" by taking the "total costs" and dividing it by the "total episodes" on the cost report.
    f. Cost reports that were either in the 5% of agencies with the highest margins or the 5% of agencies with the lowest margins.[14]
    g. For each cost report, the log of the ratio of "total payments" to "total costs" was created. Cost reports were flagged if the log of the ratio exceeded the 90th percentile of the distribution or if it was less than the 10th percentile.
3. Cost reports with any of the above flags were then dropped.
4. For each cost report, an outlier ratio was calculated by dividing the outlier payments by "total payments". If the ratio was equal to 0.10 or higher, the cost report was dropped.
5. Next, cost per visit variables[15] for each cost report were set to missing if they were in the top or bottom 1% of values for that particular variable. For example, if a cost report had a value for cost per visits for skilled nursing that was in the top or bottom 1% of the distribution of cost per visits for skilled nursing, that value would be set to missing. Agencies could have between 0 and 6 values set to missing. If any agency had all 6 values set to missing, that cost report was completely dropped from the sample.
6. Applying the above steps produces a sample of 4,313 cost reports.
    a. This number of cost reports is then reduced to 4,310 after dropping any cost report that did not have their state location listed in the provider of services file.

[12] Total payments were defined by adding together these variables: d_c*_12 d_c*_1201 d_c*_1202 d_c*_1203 d_c*_1204 d_c*_1205 d_c*_1206 d_c*_1207 d_c*_1208 d_c*_1209 d_c*_1210 d_c*_1211

[13] Total costs were defined by adding together these variables: c_3_c9_16 c_4_c6_19

[I] In this step, margins are calculated by taking ("total payments" – "total costs")/"total payments". Total payments are defined in the same way as in footnote 1, with the addition of d_c1_11. "Total costs" are defined in the same way was in footnote 2

[15] These variables included: c_1_c4_1 c_1_c4_2 c_1_c4_3 c_1_c4_4 c_1_c4_5 c_1_c4_6

b. From that sample, an additional provider was dropped because it contained no claims where visits were recorded for all six disciplines.
c. The final sample contained 4,309 cost reports.

Then payments and costs were assigned to episodes using the following approach

1. Claims from the sample in 2009 were merged to the sample of cost reports described above based on the last date of the episode. If the last date fell on or between the cost report's fiscal year start or end date, that claim was included in the sample. This was done both for a 100% sample of data from 2009 (to examine profitability under the current payment system) and a 20% sample of data from 2009 (to examine profitability under the revised payment weights and other changes to payment that go into the CY 2012 home health payment rule).
2. For the current payment system, the episode's payment equaled the payment that was on the claim minus the NRS payment for the episode. For the revised payment weights, the episode payment equaled what would be paid under the proposed payment changes for CY 2012 (a 5.06 payment reduction, a wage index update, and a home health payment update). The NRS payment was not included in this simulated payment.
3. The cost of each episode was defined using the number of visits on the claims and the cost per visit information that was on the cost report. Since some agencies had missing cost per visit information due to the 1% incremental trim, any missing cost per visit values were imputed by using the average of all non-missing values in the sample. Then, the cost of each episode was defined by multiplying each agency's cost per visit information by the number of visits on each claim for that agency. This was done separately for each discipline. For each claim, the costs computed from that multiplication were added across each discipline to construct an overall cost of the episode. The costs for NRS were also excluded in our calculations of the costs of episodes.

Using the 2008 refinement weights and the 100% sample of claims from 2009, we found that normal episodes, which are defined as non-LUPA, non-outlier, and non-PEP episodes, with between 14 and 19 therapy visits had on average payments that exceeded costs by $1,113. For normal episodes with 20 or more therapy visits, the corresponding number was $1,538. Using the 2007 revision weights and other payment changes that were proposed for CY2012 (a 5.06 payment reduction, a wage index update, and a home health payment update), we found that there was a more even payment amount exceeding costs for episodes. In addition, normal episodes providing up to 25 visits (which includes over 97% of all normal episodes) had payments that exceeded costs. This analysis shows that even after the revision of the weights and other proposed changes to the payment system for CY2012 (a 5.06 payment reduction, a wage index update, and a home health payment update), episodes with a high level of therapy would remain profitable.

3.3.4 Predictive Ratios by Payment Groups

Tables 8a and 8b present predictive cost ratios for each of the 153 payment groups. We compute the ratios in the same manner as we did in Tables 6a and 6b and then group the episodes into the case mix groups (i.e., home health resource groups [HHRGs]). Table 8a applies the three sets of weights that were used in Tables 6a and 6b to a 20% sample of normal episodes from 2005. Table 8b applies the same three sets of weights to a 20% sample of normal episodes from 2008.

In Table 8a, 55 of the payment groups had a predictive ratio above one when using the weights from the 2008 refinements. Only 28 of the payment groups had a predictive ratio above one when using the revision of the case-mix weights using episodes from 2005. Similarly, only 27 of the payment groups had a predictive ratio above one when using the revision of the case-mix weights using episodes from 2007. A similar pattern is seen in Table 8b. 44 of the payment groups in that table had a predictive ratio above one when using the weights from the 2008 refinements. Only 27 of the payment groups had a predictive ratio above one when using either of the revisions of the case mix weights using episodes from 2005 or 2007. This outcome fits in with the interpolation. The payment groups with fewer therapy visits (particularly 0 to 5 visits) had their weights adjusted upward, which should increase their predictive ratio. The groups with a lower number of therapy visits comprise a larger portion of all episodes so there are fewer payment groups with a ratio above one. The weights for the payment groups with a high number of therapy visits were dropped, which decreases those predictive ratios.

3.4 Regression of Wage Weighted Minutes of Care on Final Weights

Table 9 provides results from a linear regression where the dependent variable equals an episode's wage weighted resource minutes and the only independent variable is the episode's final payment weight when applying the 2008 refinement weights, the revision of the case mix weights using episodes from 2005, or the revision of the case mix weights using episodes from 2005. We estimate nine models in total, using each of those three sets of payment weights on a 20% sample of normal episodes from 2005, 2008, or 2009. The regression is used to determine how well the final payment weights correlate with an episode's wage weighted resource minutes.

Table 9 shows the results from each of the weighting variations are somewhat similar. For example, when looking at the regression results when using the episodes from 2005, the adjusted R-squared ranges from 0.4386 to 0.4505 depending on which set of weights is used as the independent variable. The weights from the 2008 refinements have a slightly larger adjusted R-squared value for each year of episodes used in estimating the regression. This likely occurs because we made adjustments to the weights so that low therapy payment groups had higher weights and high therapy payment groups had lower weights.

The table also shows the weights perform better using more recent years of episodes. This corresponds to our findings in Table 3 that show the adjusted R-squared of the payment regression is increasing over time. In more recent years, the payment regression variables have done a better job of predicting wage weighted resource minutes than in the past.

4 Discussion

The revision of the case-mix weights was designed to produce a new set of payment weights to improve upon those generated during the 2008 refinements. There are three major differences between the revised weights and the weights from the 2008 refinements:

(1) The version of the payment weights we use comes from a payment regression estimated using a 20% sample of normal episodes from 2007 instead of 2005.

(2) The refinement weights do not include in the hypertension diagnosis group two ICD-9 codes related to hypertension, 401.1 and 401.9. In recent years, we have found that these codes are not correlated with wage weighted minutes of care and therefore should not enter into the payment regression.

(3) The refinement weights are not generated using a payment regression with indicator variables for different therapy visit ranges. One criticism of the weights generated from the 2008 refinements was that the payment regression used to generate those weights relied too heavily on therapy visits. Therefore, we have removed those variables from the revision of the case-mix weights methodology.

While producing the methodology for the revised weights, we tested several variations on the process. These included:

(1) Although not discussed in this report, we began with an approach to revising the case-mix weights that was similar to the approach used in the 2008 refinements but which was more aggressive in pushing down the weights as the number of therapy visits increased. However, this approach produced weights for certain high therapy payment groups that were higher than the weights from the 2008 refinements. Also, the stability of the weights was poor as episodes increased in therapy visits. There were several inversions where increasing an episode's number of therapy visits (while holding all else constant) would lower the episode's weight. These outcomes led us to change the methodology to the one discussed in this report.

(2) Using the four equation model from the 2008 refinements versus a four equation model where the 401.1 and 401.9 codes were not included in the hypertension diagnosis group.

(3) Using a payment regression model that included the 401.1 and 401.9 codes versus a payment regression model where the 401.1 and 401.9 codes were dropped.

(4) Using a payment regression model that was estimated on episodes from 2005 versus episodes from 2008.

(5) Using a payment regression model which contained the same variable list as was used in the 2008 refinements versus a payment regression model where the therapy visit range indicator were dropped.

(6) Adjusting the raw payment weights by a variety of adjustment factors in order to increase the weights of episodes in payment groups with a low number of therapy visits and to decrease the weights of episodes in payment groups with a high number of therapy visits.[16]

[16] As part of this work, we multiplied the raw weights associated with therapy visit range 1 by 1.375 and multiplied the raw weights from therapy visit range 6 through therapy visit range 8 by 0.975 and multiplied

The revised weights were adjusted so that the final weights would produce the same level of payments (not including outlier payments) as the weights from the 2008 refinements when applied to a 20% sample of episodes from 2009. However, even though the weights are budget neutral overall, differences do arise in the revised weights versus the 2008 refinement weights. For example,

(1) Within a particular combination of whether the episode is early or late, its clinical level, and its functional level, the revised weights increase at a much slower rate as an episode's number of therapy visits increase compared to the 2008 refinements.

(2) Weights for payment groups with a low number of therapy visits are higher under the revised weights compared to the 2008 refinement weights. Correspondingly, weights for payment groups with a high number of therapy visits are lower under the revised weights compared to the 2008 refinement weights.

(3) Predictive ratios generated using the 2008 refinement weights and the revised weights have two main differences:

   a. Because of the upward and downward adjustments, the range of ratios for the revision weights is larger than the range of ratios for the 2008 refinement weights.

   b. In some cases the ratios from the revised weights are smoother when compared to the refinement weights. For example, the ratios from the revised weights are more likely to start out high for the lowest decile of episodes (when episodes are arrayed according to payment weights) and end low for the highest weight decile.

In some cases the ratios from the revised weights are smoother when compared to the refinement weights. For example, the ratios from the revised weights are more likely to start out high for the lowest weight decile of episodes (when episodes are arrayed according to payment weights) and end low for the highest weight decile.

the raw weights from therapy visit range 9 by 0.950. We tested two other variations on these adjustments (1) Multiply the raw weights associated with therapy visit range 1 by 1.15. Multiply the raw weights from therapy visit range 6–8 by 0.90. Multiply the raw weights from therapy visit range 9 by 0.80. (2) Multiply the raw weights associated with therapy visit range 1 by 1.075. Multiply the raw weights from therapy visit range 6–8 by 0.950. Multiply the raw weights from therapy visit range 9 by 0.900.