1-2010

Oasis Inter-Rater Reliability and Reimbursement: A Study of Inter-rater Reliability of the Outcome and Assessment Information Set (OASIS): Its Effects on the Home Health Resource Group (HHRG) and Reimbursement

Paul A. Shew

Oasis Inter-Rater Reliability and Reimbursement

A Study of Inter-rater Reliability of the Outcome and Assessment Information Set (OASIS): Its Effects on the Home Health Resource Group (HHRG) and Reimbursement

PAUL A. SHEW, PT, DPT, SHERYL L. SANDERS, PhD, NOELLE C. ARTHUR, PT, DPT, AND KENNETH W. BUSH, PhD, PT

One of the outcomes of the initial Outcome and Assessment Information Set (OASIS) is to establish a level of reimbursement through the Medicare Prospective Payment System (PPS). Several inter-rater reliability (IRR) studies have noted differences in responses to individual questions within the OASIS but little is known about how variations in scoring might affect the final reimbursement projection. In a one-way repeated measures design study, the OASIS was completed on 52 patients by both an RN and a PT within a 24-hour timeframe and the projected reimbursement rates were compared. Fifty-four percent of the outcomes of the assessment pairs were identical; differences in the remaining 46% were equally divided between RNs and PTs in projecting a higher reimbursement rate and with similar distributions. The mean difference in projected reimbursement rates was $16.43 per episode with no significant difference between the PT and RN Home Health Resource Group (HHRG) distributions.

The **O**utcome and **A**ssessment **I**nformation **S**et (OASIS) has been in use since October 2000 and is mandated to be completed for all patients admitted to Home Health Agencies (HHA) who have Medicare or Medicare managed insurance. The OASIS, a data collection tool, is a set of questions and observations made by the admitting team member. This team member may be a RN, a physical therapist (PT), or a speech language pathologist. Differences in individual raters' evaluation methods, interpretation of tests, the individual's skill level, discipline, and the understanding of each OASIS item may result in differences in responses which could result in inconsistent final scores.

One of the outcomes of the initial OASIS data collection tool is to establish a level of reimbursement through the Medicare Prospective Pay System (PPS). Teenier (2008) outlined the latest changes in calculating the way Centers for Medicare and Medicaid Services reimburse the HHA through the PPS. The results of the information collection classify the patient admission into a case-mix group based on a Home Health Resource Group (HHRG) and a dimension based on the episode and projected rehabilitation needs. There are 153 case-mix group categories defined by the Home Health PPS.

The HHRG score is established during the initial assessment by the clinician answering the uniform set of OASIS questions based on the patient's clinical severity, functional status, and service utilization. HHRG scoring is the result of the clinicians' responses to a range of OASIS items that contribute to additional HHRG points. One example is related to wounds. A stage 1 or 2 pressure ulcer is given 15 points in the Clinical (C) domain, whereas a stage 3 or 4 pressure ulcer is given 36 points. From the rating points a total number of points is computed for each of the three domains. Each of the domains is then assigned a rating of 1 through 3 (minimal, moderate, or high acuity) for final "Clinical" and

"Functional"(C, F) scores and 1 through 5 for the "Service" (S) score. The final HHRG is expressed with a CFS rating, for example C1F1S3 or C2F3S2. The case dimension is based on the episode and projected rehabilitation utilization. The HHRG score is then combined with the dimension to determine the case-mix group and final projected reimbursement rate for the 60-day episode.

Madigan et al. (2003) expressed concerns regarding inter-rater reliability (IRR) or consistency of the judgments made by two raters. Portney and Watkins (2000) frame the issue: "Reliability is fundamental to measurement because without it, we cannot have confidence in the data we collect or the conclusions we draw from those data" (p. 61). Considering the case of OASIS in Home Healthcare IRR is of interest both in the evaluation of specific items within the assessment as well as, how inconsistencies affect final reimbursement rates.

Prior studies have evaluated IRR regarding specific OASIS items. To assess this, the Kappa coefficient was most commonly used. According to Landis and Koch (1977) the thresholds for levels of agreement for the Kappa coefficient are: 0.00 to 0.20 = poor, 0.21 to 0.40 = fair, 0.41 to 0.60 moderate, 0.61 to 0.80 = substantial and above 0.80 almost perfect. Hittle et al. (2003) and Madigan and Fortinsky (2005) considered correlations lower than 0.6 Kappa coefficient to be below an acceptable threshold for IRR regarding the OASIS.

Hittle et al. (2003) performed a sequential study involving RNs who were hired, trained, and tested specifically in order to minimize the individual variation in assessment skills when administering the OASIS. The researchers reported that in the 25 cases studied, a large number of OASIS items demonstrated a very high correlation of reliability. However, 16 of the 92 (17%) primary items had an agreement score of less than 0.60-weighted Kappa with the greatest discrepancies noted in the Functional Domain.

In another study, Madigan & Fortinsky (2003) measured the accuracy of raters completing the OASIS. In this study, a videotape was made of a simulated complex home health assessment that included OASIS items. The video was reviewed by four expert home health nurses who developed an answer key of "correct answers." The video was then shown to 436 raters (nurses, PTs, occupational therapists, and speech-language pathologists) from 29 Ohio HHA, with each rater scoring the OASIS based on their observations from the video. The accuracy was found to be above 0.80 correct for 11 of the 19 (58%) items evaluated. In 42% of the items evaluated accuracy was below 0.60. A comparison of nurses and therapists demonstrated a high correlation in 10 of the 16 items evaluated and significant differences in 6 of 16 items. This study also reported on a degree of built in tolerance related to the HHRG system suggesting that several numeric values within each domain may allow for the same final HHRG score.

Kinatukara et al. (2005) reported on IRR of two nurses. The first was an experienced nurse who was hired specifically for the study (research clinician) and the second was a staff nurse clinician from the agency. This study utilized two methodological approaches. The first approach involved 259 patients and used sequential assessments. Both evaluators completed the OASIS with a delay of 24 to 72 hours between assessments. The IRR was shown to be less than 0.60 (moderate) for 55 of the 60 items studied and less than 0.40 (fair) for 39 of the test items. With the second method in this research the two raters made simultaneous OASIS assessments. The staff clinician was responsible for all aspects of the OASIS evaluation while the research clinician acted only as an observer during the completion of the second OASIS. This study reported that 65% of the OASIS items studied demonstrated IRR less than 0.60 and 29% demonstrated fair or less than fair reliability (less than 0.4).

Madigan et al. (2005) used the simultaneous approach to collect OASIS data by home health field staff from 88 patients. During the selected visits, the first rater completed the OASIS as part of the agency admission protocol and the second rater concurrently completed the second OASIS while observing the patient and the first rater. For this study 25 OASIS items were evaluated with two (8%) having a weighted kappa coefficient below 0.60.

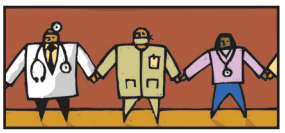
Neal (2000) conducted a study to demonstrate IRR between RNs and also between PTs and RNs. For this research, one of two study raters was

used to complete a second OASIS assessment within 24 hours of the agency admission OASIS. A complete list of all OASIS items evaluated was not provided in the publication; however, the overall IRR results were reported using another reliability coefficient. The agency OASIS had been completed by either an RN (n = 14) or a PT (n = 9). The overall percent agreement for study RN to staff RN was 0.65 and for study RN to staff PT was 0.60.

In an unpublished study involving 20 subjects, Arthur (2007) compared OASIS IRR of a Nurse to a PT. For this research the primary clinician was a field RN who completed the patients' initial OASIS as a portion of the full admitting assessment. The PT completed an abbreviated OASIS assessment within 36 hours. This study demonstrated that the IRR of 67% of the OASIS items was below 0.60 and 52% were below 0.40 using the weighted Kappa coefficient.

Each of the above studies demonstrated some level of inconsistency in scoring selected items within the OASIS. Differences appear to be greater in HHA field clinicians as compared to staff hired and trained specifically for a study. In addition, Madigan et al. (2003), Madigan and Fortinsky (2005), and Arthur (2007) all noted a significant variance between scores provided by nursing staff as compared to rehabilitation staff. These differences have the potential of resulting in variations in reimbursement rate for the agency since the final HHRG is based on responses to individual items.

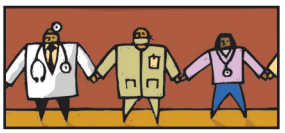
In contrast to prior studies, which evaluate agreement of each test item, the purpose of this study was to compare the final scores of the OASIS between home health professionals with regards to reimbursement rate, as well as ratings of clinical and functional domain scores for HHRG. The difference of projected reimbursement rates may be great enough for agencies to determine the level of continuing education required by the agency for individual clinicians to complete the OASIS more consistently. Significant differences in the final Clinical and Functional domain scores may give information on where further continuing education for field staff needs to focus. It was hypothesized that a significant difference in reimbursement rates occurs following admission to OASIS assessment by nursing staff compared to rehabilitation staff.



The purpose of this study was to compare the final scores of the OASIS between home health professionals with regards to reimbursement rate, as well as ratings of clinical and functional domain scores for HHRG.

Methods

This study was approved by the Institutional Review Board at Pacific University in Forest Grove, Oregon and approved by the HHA. The agency is a nonprofit Medicare certified HHA serving urban, suburban, and rural patients. The data collection took place during the course of home health clinicians' regular work day in order to include as many factors that may affect the outcome as possible including case load, time of day, scheduling, or stresses of the many details that a home healthcare clinician experiences that may affect his or her decision-making process at any point in time. A sequential design was employed in order to limit the influence between clinicians completing an OASIS, as might occur during a simultaneous visit, and the choice of the 24-hour time frame was used to minimize the potential change in patient condition between OASIS assessments. If longer than 24 hours the patient may improve materially and become less likely to reflect the conditions found by the first rater. This was accomplished by the second rater noting the time of the first rater's visit by reviewing the scheduling timestamp on their laptop computer which was used during the course of the second



With an error probability of 5%, it can be said from this small study, that using the Wilcoxon signed-rank test, there is no statistically significant difference between the dollar value of RN and PT ratings.

rater's visit for the non-OASIS portion of their visit.

Home Health patient participants were selected by the home health schedulers based on the potential for an admission assessment to be completed by an agency RN and a PT within 24 hours. The patient participants were informed of the study and a consent form was signed either by the patient or legal representative. The participants (n = 52) were over 18 years old and all were receiving benefits from Medicare or Medicare-based Health Maintenance Organizations. HHA staff participants who completed the OASIS included 18 RNs and 12 PTs. All except one RN and one PT had greater than 1-year experience completing the OASIS. Within this agency four RNs and three PTs had been completing the OASIS since 2000. The home health schedulers assigned a unique patient participant number in order to keep confidential the identity of each patient. A RN (Evaluator #1) completed the initial assessment, which included the OASIS, on a laptop computer. The next day a supervisor was able to note the HHRG from that assessment by reviewing the computed results made from the OASIS of Evaluator #1. A PT (Evaluator #2), completed the second OASIS on paper during the course of their initial PT assessment and within 24 hours of the nursing OASIS assessment. The PT did not have access to the original OASIS assessment. This paper OASIS assessment was given to the principal investigator who entered the data into a separate computerized "test" patient file in order to avoid confusion with normal HHA billing. This test file was used to compute the second OASIS HHRG score. The reimbursement rate was extrapolated from the case-mix groups, average costs, and case-mix weight table. This study compared the projected reimbursement rates and HHRG scores from the RN assessment to that of the PT assessment, so a measure of the distribution of rates was found to be appropriate.

Results

Evaluations were completed on 52 pairs of OASIS data set results. For all the 52 patients in this sample the RNs predicted a total reimbursement of $124,637.14 and the PTs predicted a reimbursement of $123,782.89. The mean projected reimbursement rate for the RN was $2396.87, and for the PT was $2380.44, with a mean difference of $16.43 per episode.

A side-by-side comparison of the cases studied, 28 (54%) had HHRG scores and reimbursement rates that were equal between RNs and PTs. Of the nonequal scores five reimbursement rate projections were separated by less than 10%. Of these, the RN predicted reimbursement rate higher than the PT in three of the five pairings, and the PT predicted higher in two. There were 10 reimbursement rate projections that differed between 10% and 20% with the RN predicting higher in six cases and the PT predicting higher in four. A difference of 20% and 30% was found in nine cases with the RN predicting higher in three cases and the PT predicting higher in six. Overall the RNs scored higher in 13 of the 24 nonequal cases. These nonequal scores were compared to evaluate variations in distributions of reimbursement predictions of RN and PT to analyze if there was a significant difference in prediction of a greater reimbursement rate by the RN or PT. With an error probability of 5%, it can be said from this small study, that using the Wilcoxon signed-rank test, there is no statistically significant difference between the dollar value of RN and PT ratings.

Finally, the differences in Clinical and Functional domain HHRG scores between RNs and

PTs were compared to determine if one discipline consistently rated higher than the other in either of these categories. A consistent difference in scores between disciplines was not found. Interestingly, in this study, eight Clinical (C) scores were greater for the RN and eight were greater for the PT. The Functional (F) scores of the RN predicted eight patients at a higher rating and the PTs predicted four at a higher rating. The predicted functional scores were reviewed using the binomial probability distribution table (A-9) from Portney and Watkins. There is no statistically significant difference in the comparison among the Functional domain totals.

Conclusions

The results of this study indicate that, on the whole, clinician IRR is adequate in determining overall HHRG and reimbursement rates. Although some differences in HHRG and projected reimbursement rates were noted, 54% of the scores were the same for RNs and PTs. Of those scores that were not the same there was no indication that one discipline consistently rated patients in higher categories.

The number of projected rehabilitation visits (MO 826) was not used because reimbursement is based upon the total number of rehabilitation visits actually made which is determined at the end of the 60-day episode. Because of this the service utilization domain was not included in this study.

The team members participating in this study had been instructed not to discuss the specific OASIS results prior to the completion of the second assessment; however, it is possible that, during normal daily coordination of care exchanges some information may have been inadvertently passed. In this agency, there are regular, quarterly reviews with training for completion of the OASIS as well as an RN Quality Review Supervisor (QRS) employed by the agency, who counsels clinicians regarding specific OASIS items. For the current study, these OASIS evaluations were computed prior to the QRS review; however, it should be noted that this constant review process may affect daily practice.

This study represents only one evaluation of the comparison of OASIS results of a relatively small number of home health patients and RN and PT assessments within a single branch of one HHA and the results may not be widely generalized. Future studies could increase the number of clinician participants, home health patient participants and could involve multiple HHAs and geographic areas before conclusions can be drawn with regard to RN versus PT OASIS assessments.

The outcome for the final HHRG score may be either a result of variance built into the computation of the HHRG or of similarities in responses to individual test items. In reviewing past studies it has been noted that at least some variation of responses occurs between individual reviewers completing the OASIS and may be even greater between disciplines completing the OASIS. This study did not determine how closely the individual responses to each OASIS item were related nor did it evaluate accuracy. It might be helpful in understanding the dynamic of IRR as it affects the final outcome or reimbursement by correlating agreement of specific OASIS test items to the final reimbursement rate. This may give an indication which items either may tend to be more or less reliable between raters and disciplines or which test items may cause the greatest difference in final projected reimbursement rate.

The findings of this study support the concept of consistency of measurement between the rater disciplines of physical therapy and nursing with the dollar estimate differences of less than 1%. Some HHAs have chosen to have their RNs do all of the agency OASIS assessments for a variety of reasons including consistency of reimbursement or a more efficient utilization of PTs due to the time required to complete the OASIS. Although this study was of limited scope it does support the practice, from a reimbursement standpoint, of a HHA using either RNs or PTs for the initial OASIS assessment. ⌂

Paul A. Shew, DPT, *is a Physical Therapist, Providence Home Services Portland West, Portland, Oregon.*

Sheryl L. Sanders, PhD, *is an Associate Professor of Anatomy, School of Physical Therapy, Pacific University, Hillsboro, Oregon.*

Noelle Arthur, DPT, *is a Physical Therapist, Acute Therapies, Fletcher Allen Health Care, Burlington, Vermont.*

Kenneth W. Bush, PhD, *is a Professor of Physical Therapy, School of Physical Therapy, Pacific University, Hillsboro, Oregon.*

Address for correspondence: Paul A. Shew, PT, DPT, PO Box 697, Banks, OR 97106 (e-mail ptpapa@hotmail.com).

REFERENCES

Arthur, N. C. (2007). *OASIS*: *Study of interdisciplinary inter-rater reliability*. Doctoral Research for Pacific University School of Physical Therapy.

Centers for Medicare and Medicaid Services. (2007). *Overview*. http://www.cms.hhs.gov/OASIS/02.

Centers for Medicare and Medicaid Services. (2007). Table 5 Case Mix Groups, average costs, and case-mix weight. http://edocket.access.gpo.gov/2007/pdf/07-4184.pdf pages 49, 828-49, 832.

Hittle, D. F., Shaughnessy, P. W., Crisler, K.S., Powell, M. C., Richard, A. A., Conway, K. S., et al. (2003). A study of reliability and burden of home health assessment using OASIS. *Home Health Services Quarterly, 22*(4), 43-63.

Kinatukara, S., Rosati, R., & Huang, L. (2005). Assessment of OASIS reliability and validity using several methodological approaches. *Home Healthcare Services Quarterly, 24*(3), 23-37.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Madigan, E. A. (2002). The scientific dimensions of OASIS for home care outcome measurement. *Home Healthcare Nurse, 20*(9), 579-583.

Madigan, E. A., Tullai-McGuinness, S., & Fortinsky, R. H. (2003). Accuracy in the outcomes and assessment information set (OASIS): Results of a video simulation. *Research in Nursing and Health, 26*, 273-283.

Madigan, E. A., & Fortinsky, R. (2005). Inter-rater reliability of the outcomes and assessment information set (OASIS): Results from the field. *Gerontologist, 44*(5), 689-692.

Neal, L. J. (2000). OASIS-inter-rater reliability. *Caring Magazine, 19*(8), 44-47.

Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research applications to practice*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Teenier P. (2008). 2008 refinements to medicare home health prospective payment system. *Home Healthcare Nurse, 26*(3), 181-184.