Overview of Risk Adjustment and Outcome Measures for Home Health Agency OBQI Reports: Highlights of Current Approaches and Outline of Planned Enhancements[1, 2]

by

Peter W. Shaughnessy, PhD and David F. Hittle, PhD
Center for Health Services Research, University of Colorado Health Sciences Center

September 2002

The purposes of this document are (1) to briefly discuss outcome measurement, risk adjustment, and the rationale for risk adjusted outcome reporting in the context of outcome-based quality improvement (OBQI) for home health care, and (2) to summarize the current risk adjustment and outcome measurement approaches as well as plans for their continued refinement. Except for Section 1 and the first component of Section 2, this overview is intended for a somewhat technically inclined audience. Nonetheless, it should be useful to those with little or no research or statistical experience. For such individuals, an intuitive sense of the risk adjustment methods can be acquired by simply reading the narrative and skimming through the material that is more technical in nature. In Sections 5 and 6, background information on outcome measures and OBQI is presented in order to provide contextual information that is relevant to understanding the applications of the statistical risk adjustment methods summarized here.

1. What Risk Adjustment Is and Why It Is Needed

Suppose the hospitalization rate is 20% for one home health agency and 30% for another. On the basis of these statistics alone, one might conclude that the second agency provides inferior care because its patients are hospitalized more often. However, if the case mix of the second agency is radically different from that of the first agency, such a conclusion could be invalid. For example, suppose the average age of the second agency's patients is 15 years older than the first agency, and it has considerably more patients with severe disabilities. In this case it might be understandable or expected that its hospitalization rate would be higher. The various patient-level factors that influence (positively or negatively) the likelihood of hospitalization are termed *risk factors* for hospitalization. The purpose of *risk adjustment* when comparing outcome rates (e.g., hospitalization rates) for two different patient samples is to statistically compensate (or adjust) for risk factor differences in the two samples so that the outcome rates can be compared legitimately despite the differences in risk factors.

Conceptually, it is possible to enumerate a large number of risk factors that *might* influence a given outcome. Practically speaking, however, each outcome measure used in producing risk-adjusted outcome reports for OBQI tends to have a limited number of risk factors (from 20 to about 50) available from the Outcome and Assessment Information Set (OASIS) that can be empirically determined to exert a substantial impact on that outcome. In general, risk factors for an outcome are chosen first by conceptually and clinically specifying the potential risk factors, and then assessing which ones are empirically related to the outcome.

---

[1] A detailed description of the demonstration programs and research methods that are referenced in various places in this summary document can be found in the four-volume report series entitled "OASIS and Outcome-Based Quality Improvement in Home Health Care: Research and Demonstration Findings, Policy Implications, and Considerations for Future Change" by Shaughnessy, Crisler, Hittle, et al., Denver, CO: Center for Health Services Research, February 2002. See further the reference section at the end of this paper. The four volumes are available on www.cms.hhs.gov/providers/hha.

[2] As research proceeds, the material presented here will be augmented with further specifics that address a variety of quality/measurement, risk adjustment, attribution-of-effects, sample size, statistical testing, unit-of-analysis, and empirical validation issues that are beyond the scope of this summary document.

2. Risk Adjustment Methodology

For purposes of discussion, assume that for a given year the outcomes of patients discharged from Home Health Agency A are to be compared with the outcomes of patients from all home health agencies throughout the United States -- either by selecting a sample from all patients nationally for the given year or actually using all home care patients throughout the U.S. as the reference group. In this case, refer to the patients from Home Health Agency A as the "test group" and those from the nation as the "comparison group."[3]

Statistical Modeling: One of the most straightforward ways to risk adjust an outcome in order to compare Agency A with agencies from the rest of the United States is to produce an expected value for Agency A's outcome based on the relationship between the outcome and its risk factors as this relationship exists in the comparison group (i.e., the national sample). For example, by analyzing and estimating the empirical relationship between improvement in ambulation and its risk factors in the U.S. population of home health patients, one can develop a formula expressing this outcome as a mathematical function of the risk factors. Using this formula for each of Agency A's patients, it is possible to calculate an expected outcome rate for Agency A (for all patients dependent in ambulation) under the assumption that the relationship between improvement in ambulation and its risk factors is the same for Agency A's patients as it is for home care patients in the rest of the country. If the expected outcome rate for Agency A is lower than the actual outcome rate for Agency A's patients, then Agency A would be considered above average on this particular outcome. Conversely, if it were higher, then Agency A would be considered below average. Furthermore, it is possible to quantify the magnitudes of the differences between observed and expected outcome rates, and report their statistical significance.

Risk Adjustment Methods To Date: Statistical modeling provides a means to estimate the relationship between an outcome and a set of risk factors. There are a variety of ways to estimate statistical models that can be used as predictive formulas expressing an outcome as a function of multiple risk factors (Iezzoni, 1994). In the research work that led to the development of the OBQI program as currently structured, several alternative methods were tested. These included but were not limited to logistic regression, classification and regression tree (CART) methods, the general method of data handling (GMDH), and statistical standardization with strata. After assessing the utility of the various methodologies (including combinations that involved more than one method), logistic regression was selected to produce risk-adjusted outcome reports for home health agencies in the OBQI demonstration programs. The approach to using logistic regression is summarized in the remainder of this section. Some of the refinements and further research on methods that will be undertaken over the next three years are outlined in Section 7.

Overview: Logistic Regression as Used in the National and New York State OBQI Demonstration Trials:[4] Risk adjustment of outcome measures used in the OBQI demonstrations was based on logistic regression models that were estimated for each outcome using the entire pool of patients (the reference or comparison group of patients) from all agencies participating in the

---

[3] Owing to the fact that certain types of patients are excluded for different outcomes, the test groups from Agency A (and therefore the analogous comparison groups from the U.S. population) tend to differ for each outcome under consideration. For example, the Agency A patient sample and the comparison patient sample used to compare risk-adjusted outcome rates for improvement in ambulation are restricted only to those patients who are dependent in ambulation.

[4] Results of the demonstration trials are available in Shaughnessy, Hittle, Crisler, et al., 2002.

demonstrations for a given time period.[5] Models for every outcome measure then were used to obtain predicted values for each outcome for every agency. The logistic regression models also provided the means to compare an agency's outcomes for a given year with a prior year, adjusting for changes in case mix. For each of 41 outcomes, the risk-adjusted outcome report that was distributed annually for OBQI purposes in the demonstration programs presents a graphical comparison of an agency's actual outcome (1) with its expected outcome (using a "national" comparison group) and (2) with its risk-adjusted outcome for the prior year. An enumeration of the outcome measures with a brief rationale for why the current group of 41 outcomes was selected is provided in Section 5 of this document. An excerpt from a risk-adjusted outcome report and an explanation of how the models are used is presented in Section 6.

Summary of the Modeling Process: Thus, for each such outcome, a separate logistic regression model was estimated. For validation purposes, the entire pool of reference patients was randomly split into two groups, a "developmental sample" and a "validation sample" (Harrel, Lee, and Mark, 1996). A logistic regression model was estimated for each outcome using exclusively cases from the developmental sample. This process first entailed conceptually and clinically specifying the more important risk factors that were expected to influence the outcome under consideration and that could be computed (from OASIS data). About 150 potential risk factors were candidates for each risk adjustment model used in the demonstrations. The specified risk factors then were screened to determine those that were empirically related to the outcome by analyzing the statistical associations (correlations) between risk factors and the outcome (Mickey and Greenland, 1989). Then logistic regression analysis using stepwise variable selection was conducted to develop a preliminary risk adjustment model (Lee and Koval, 1997).

A series of iterative steps followed in which a logistic regression model was estimated for the outcome under consideration. Coefficients and odds ratios for each risk factor were examined to determine if they were clinically plausible and statistically reasonable (Freedman, 1983). At this stage, clinical plausibility and conceptual meaningfulness of the relationships between outcome measures and risk factors, and of the clinical/conceptual relationships among the risk factors as a group, received more attention than statistical considerations, which served more as guideposts or criteria to use to ensure pragmatically useful models. Risk factors with clinically questionable coefficients were eliminated, and often were replaced by other more meaningful risk factors or combinations of risk factors (Flack and Chang, 1987). Each model was reestimated a number of times in this process to reach a clinically and statistically stable model (Rothman and Greenland, 1998; Harrel, Lee, and Mark, 1996). Upon completion of this first stage process, the explanatory power of the model then was tested using the aforementioned "set-aside" or validation sample. This involved selecting a series of 20 random subsamples from the validation sample, calculating a predicted outcome value for each case, and then correlating the predicted outcome value with the actual outcome value to obtain 20 R^2 statistics -- one for each of the 20 random validation subsamples. Each of the R^2s reflects the proportion of the outcome variance explained by the model in the validation subsample

[5] Strictly speaking, it is appropriate to eliminate the test patients from the comparison group when doing this. However, for OBQI applications the sample size of the test group is always miniscule relative to the sample size of the comparison group, resulting in no material difference in the estimated risk models using either approach. Consequently, for any given outcome the same risk model can be used for all agencies in the context of outcome reporting, considerably simplifying the logistics and operational features of producing outcome reports.

(C-statistics are now being used as well because under certain circumstances R^2 can be low, but a higher value for C can indicate the model is useful for risk adjustment nonetheless).[6]

In instances where there was a substantial discrepancy between the explanatory power (variance explained) for the developmental sample and the validation subsamples (in the form of either a large difference between the developmental sample R^2 and the mean of the 20 validation subsample R^2s, or an unusually large range [i.e., the difference between the minimum and maximum] in the R^2s corresponding to the 20 validation subsamples),[7] this would indicate that the model had been "over-fitted" to the developmental sample. Reestimation of the model then would be required, using the developmental sample. Essentially the same steps that were followed in the initial model refinement would be repeated, changing or eliminating the (sometimes collinear) risk factors judged to cause the overfit problem, reestimating model coefficients, then reviewing the model again, until a stable and clinically reasonable model was obtained for which the developmental explanatory power and validation explanatory power were approximately the same. At this stage of the estimation/reestimation process, equal emphasis was given to clinical/conceptual and statistical considerations because problems of overfitting often warrant devoting close attention to statistical properties of coefficients, odds ratios, and statistical interrelationships among risk factors. The 41 separate risk models for each outcome were derived in this manner and used to produce the outcome reports for the first year of the demonstration programs. Thereafter, risk models were reestimated each year to produce three rounds of annual outcome reports.

Advantages of Logistic Regression: Some of the practical reasons why the logistic regression approach has been used to date (until the results of the research in Section 7 become available) are:

a. Logistical regression results either have been superior to or the same as those derived using the other methods of risk adjustment that have been tested to date in terms of explanatory power and both clinical and statistical understandability and utility.

b. Logistic regression is a standard and widely accepted methodology for risk adjustment of dichotomous outcome measures, particularly in health care. The estimated parameters and associated statistics are well known and intuitively understandable to researchers and methodologists in the health care field. Goodness of fit and diagnostic methods for model testing and refinement are reasonably well established.

c. Very importantly, the methodology permits the use of a large number of risk factors relative to alternative methods such as standardization or stratification. Models often involve 20, 30, or even 50 or more risk factors. For OBQI, it has proven useful to have a larger number of risk factors, as long as models are stable and stand up under cross-

[6] Despite the fact R^2s naturally tend to be lower for logistic regression models for (various of) ordinary least squares models because of the binary nature of the dependent variable, the R^2 statistic has emerged as one of the two or three statistics of choice for assessing explanatory power and goodness of fit for logistic regression (Mittlböck and Schemper, 1996; Agresti, 1996). Other goodness-of-fit and diagnostic statistics or statistical approaches that have been or are presently being used for developing logistic regression models for OBQI based on national data include C-statistics (area under the ROC curve) and deciles of risk (Hosmer and Lemeshow, 2000; Osius and Rojek, 1992). Each approach has strengths that complement those of the R^2 statistic.

[7] In the demonstration research, other goodness-of-fit statistics were examined, although the R^2 statistic was typically the most useful. As noted, however, C-statistics and decile-of-risk tables are now being used in conjunction with and in a manner similar to R^2s.

validation, because home care clinicians tend to have greater confidence in such models, finding them more credible than models with relatively few risk factors.[8]

d. Efficient and well-documented statistical software is available for conducting the necessary large-scale analyses to estimate each logistic regression model using exceptionally large sample sizes, including reestimating models to keep them current from year to year.

e. Reporting software can be designed in a relatively straightforward manner to incorporate logistic regression models, and readily updated as models change from year to year.

Risk Adjustment for the Quality Improvement Organization (QIO, formerly PRO) OBQI Pilot Project:[9] The logistic regression approach also was used to produce risk-adjusted outcome reports for home health agencies participating in the QIO OBQI pilot project in five states. The models developed for purposes of outcome reporting for the OBQI demonstrations were reestimated and revalidated using data from the CMS OASIS national repository. The national repository data were more recent than the demonstration data and more representative of the entire nation. In addition, the number of cases available for analysis was substantially greater. A similar methodology to that described above for estimating risk models was followed in this OBQI pilot project.

3. Potential Risk Factors Included in Model Development Process

As noted earlier, a total of approximately 150 measures have been used as potential risk factors in the risk adjustment process (Table 1) in the OBQI demonstrations as well as the QIO pilot and the first round of national OBQI outcome reports in early 2002. The risk factors are computed from OASIS data collected at start or resumption of care and therefore represent baseline patient status for each episode of care. Length of stay categories are exceptions to this rule.[10] All of the risk factors listed in Table 1 are considered as candidates for inclusion in each outcome measure's risk model. As indicated above, specific risk factors are selected for each outcome based on both clinical and statistical criteria. Therefore, the actual risk factors included in risk adjustment models differ from outcome to outcome.

---

[8] Since sample sizes will ultimately be in the millions, certain statistical considerations (such as the ratio of sample size to the number of risk factors) and model instability or the tendency to overfit models to data because this ratio might be low) are perforce less problematic. In fact, if multicollinearity and clinical meaningfulness are properly addressed (Agresti, 1996), the validation and stability of a model estimated using extremely large samples are enhanced as the number of clinically and conceptually meaningful risk factors in a model increases.

[9] The QIO pilot project was a sequel to the National and New York State Demonstration Trials. It entailed two rounds of annual outcome reports for approximately 400 home health agencies in five states implementing OBQI with the support of QIOs in these states. This project was expanded under the QIO seventh scope of work to include other states.

[10] Length of stay poses a unique challenge in risk adjustment for home health patients. On the one hand, it is clearly tied to natural progression of disease and disability, and should therefore be used as a risk factor. On the other hand, it is correlated with the volume of service provision and therefore related to the treatment. Thus far, a middle ground position has been employed, not using length of stay as a continuous risk factor, but using instead indicator variables of broad categories of length of stay as risk factors. See Table 1 for these categories.

TABLE 1: Risk Factors Used in Outcome Risk Adjustment Process.

- Age
- Gender: female
- Any Medicare payment source
- Patient resides in own home
- *Structural barriers:
  - Stairs inside – must be used
  - Stairs inside – optional use
  - Stairs to access house
- *Infested, trashed, cluttered, soiled living area
- *Structural hazard (floor, roof, etc.)
- *No fire safety device
- Patient lives alone
- Patient lives with family member
- Patient has unpaid live-in help
- Presence of a primary caregiver
- ADL assistance provided by caregiver
- IADL assistance provided by caregiver
- Frequency of caregiver assistance
- Inpatient discharge from hospital
- Inpatient discharge from rehab. facility
- Inpatient discharge from nursing home
- Medical regimen change
- Prior condition: Urinary incontinence
- Prior condition: Catheter
- Prior condition: Intractable pain
- Prior condition: Impaired decision-making
- Prior condition: Inappropriate behavior
- Prior condition: Memory loss
- Severity rating of primary diagnosis
- Maximum diagnosis severity rating
- Number of severity ratings >= 2
- Moderate recovery prognosis
- Good rehabilitation prognosis
- High risk factors:
  - Smoking
  - Obesity
  - Alcohol
  - Drugs
- Activities of daily living (ADL) disabilities at SOC/ROC:
  - Grooming
  - Dressing upper body
  - Dressing lower body
  - Bathing
  - Toileting
  - Transferring
  - Ambulation
  - Eating
- Activities of daily living (ADL) disabilities prior to SOC/ROC:
  - Grooming
  - Dressing upper body
  - Dressing lower body
  - Bathing
  - Toileting
  - Transferring
  - Ambulation
  - Eating
- Instrumental activities of daily living (IADL) disabilities at SOC/ROC:
  - Light meal preparation
  - Transportation
  - Laundry
  - Housekeeping
  - Shopping
  - Phone use
  - Management. of oral medications
- Instrumental activities of daily living (IADL) disabilities prior to SOC/ROC:
  - Light meal preparation
  - Transportation
  - Laundry
  - Housekeeping
  - Shopping
  - Phone use
  - Management of oral medications
- Dyspnea
- Vision impairment
- Hearing impairment
- Speech/language impairment
- Pain interfering with activity
- Intractable pain
- Cognitive functioning
- Confusion frequency
- Anxiety Level
- Frequency of behavior problems
- Memory deficit
- Impaired decision-making
- Verbal disruption
- Inappropriate behavior
- *Disinterested, disturbed sleep, no appetite
- *Indecisive
- Depression scale
- Stasis ulcer(s) present
- Number of stasis ulcers
- Status of most problematic stasis ulcer

***** Items marked by an asterisk were originally used as potential risk factors in the OBQI demonstration trials and for the first round of national outcome reporting as well as the OBQI pilot. However, these items have more recently been deleted from OASIS and therefore are not available for risk adjustment. One of the reasons why it was acceptable to drop these items is that they appeared in relatively few models. However, it is possible that with increased refinement and analysis, the items could have been more consequential as risk factors.

(continued on next page)

TABLE 1: Risk Factors Used in Outcome Risk Adjustment Process. (Cont'd)

- Surgical wound(s) present
- Status of surgical wound
- Number of surgical wounds
- Pressure ulcer(s) present
- Stage of most problematic pressure ulcer
- Status of most problematic pressure ulcer
- Stage 2-4 ulcer(s) present
- Stage 3-4 ulcer(s) present
- Urinary tract infection past 14 days
- Urinary incontinence day and night
- Urinary incontinence
- Urinary incontinence or catheter
- Urinary incontinence severity scale
- Urinary catheter
- Bowel incontinence
- Ostomy
- Acute Conditions
  - Orthopedic condition
  - Neurological condition
  - Open wounds/lesions
  - Terminal condition
  - Cardiac/peripheral vascular condition
  - Pulmonary condition
  - Diabetes mellitus
  - Gastrointestinal disorder
  - Contagious/communicable disease
  - Urinary incontinence/catheter
  - Mental/emotional condition
  - Oxygen therapy
  - IV/infusion therapy
  - Enteral/parenteral nutrition
  - Ventilator
  - Other acute condition
  - Number of acute conditions
- Chronic Conditions
  - Dependence in living skills
  - Dependence in personal care
  - Impaired ambulation/mobility
  - Eating disability
  - Urinary incontinence/catheter
  - Dependence in medication administration
  - Chronic pain
  - Cognitive/mental/behavioral condition
  - Chronic condition with caregiver
  - Number of chronic conditions
- Diagnoses:
  - Infectious/parasitic diseases
  - Neoplasms
  - Endocrine/nutritional/metabolic
  - Blood diseases
  - Mental diseases
  - Nervous system diseases
  - Circulatory system diseases
  - Respiratory system diseases
  - Digestive system diseases
  - Genitourinary system diseases
  - Pregnancy problems
  - Skin/subcutaneous diseases
  - Musculoskeletal system diseases
  - Congenital anomalies
  - Ill-defined conditions
  - Fractures
  - Intracranial injury
  - Other injury
- Length of stay:
  - More than 31 days
  - More than 62 days
  - More than 124 days

4. Examples of Risk Adjustment Models

Examples of risk adjustment models are presented for two outcomes: Improvement in Dressing Lower Body (Table 2) and Acute Care Hospitalization (Table 3). For every risk factor in each model, the logistic regression coefficient is presented, along with the statistical significance associated with that coefficient, the corresponding odds ratio, and the 90% confidence interval for the odds ratio for each risk factor.[11] The variance explained (R^2) by each model is shown for the developmental sample in conjunction with summary statistics (minimum, maximum, and mean R^2) for the 20 validation subsamples for the outcome under consideration.

[11] The odds ratio is calculated as the exponential function of the logistic regression coefficient.

TABLE 2: Logistic Regression Model for Risk Adjusting the Outcome of Improvement in Dressing Lower Body.

Risk Factor Measured at SOC/ROC*†	Coefficient‡	Odds Ratio‡	(90% CI)‡
Medicare as a payment source (0-1)	.18	1.19	(1.10-1.29)
Patient lives alone (0-1)	.22	1.24	(1.15-1.34)
ADL assistance provided by caregiver (0-1)	-.22	.80	(0.76-0.86)
Inpatient discharge from hospital (0-1)	.36	1.43	(1.34-1.52)
Inpatient discharge from rehab. facility (0-1)	.28	1.32	(1.20-1.46)
Disruptive/socially inappropriate behavior within past 2 weeks (0-1)	-.26[b]	.77	(0.62-0.96)
Memory loss requiring supervision within past 2 weeks (0-1)	-.11[b]	.90	(0.81-0.99)
Disability in dressing upper body (0-3)	-.09	.92	(0.87-0.96)
Disability in dressing lower body (0-3)	.90	2.46	(2.31-2.62)
Disability in toileting (0-4)	-.10	.90	(0.87-0.94)
Disability in transferring (0-5)	-.20	.82	(0.78-0.86)
Disability in ambulation (0-5)	-.17	.84	(0.80-0.89)
Disability in housekeeping (0-4)	-.06	.94	(0.91-0.98)
Disability in light meal preparation (0-2)	-.07[a]	.93	(0.88-0.99)
Disability in mgt of oral medications (0-2)	-.19	.83	(0.79-0.87)
Disability in telephone use (0-5)	-.12	.89	(0.87-0.91)
Prior (2 weeks ago) disability in grooming (0-3)	-.08	.93	(0.89-0.97)
Prior (2 weeks ago) disability in ambulation (0-5)	-.07	.93	(0.89-0.97)
Prior (2 weeks ago) disability in laundry (0-2)	-.20	.82	(0.78-0.87)
Moderate recovery prognosis (0-1)	.17	1.19	(1.07-1.32)
Good functional status rehab prognosis (0-1)	.20	1.22	(1.13-1.31)
Stasis ulcer(s) present (0-1)	-.33	.72	(0.59-0.88)
Number of surgical wounds present (0-4)	.12	1.12	(1.08-1.17)
Status of most problematic pressure ulcer (0-3)	-.14	.87	(0.82-0.93)
Acute condition: orthopedic (0-1)	-.23	.80	(0.74-0.85)
Acute condition: ventilator (0-1)	-1.04[b]	.35	(0.13-0.97)
Acute condition: terminal (0-1)	-.35	.71	(0.60-0.84)
Chronic condition: dependence in living skills (0-1)	-.10[b]	.90	(0.82-0.99)
Chronic condition: dependence in personal care (0-1)	-.44	.64	(0.59-0.70)
Chronic condition: urinary incontinence/catheter (0-1)	-.34	.72	(0.66-0.78)
Diagnosis: ill-defined conditions (0-1)	-.08[b]	.93	(0.87-0.99)
Diagnosis: nervous system disorder (0-1)	-.17	.84	(0.77-0.92)
Length of stay: more than 31 days (0-1)	.50	1.65	(1.56-1.75)
Constant	.08		

Developmental sample R^2 = 0.230; Validation sample R^2s for 20 validation subsamples: Minimum = 0.211, Maximum = 0.237, and Mean = 0.224.§

* SOC = Start of Care, ROC = Resumption of Care after inpatient stay. Risk factors pertain to SOC/ROC unless indicated otherwise.

† The number of values in the measurement scale for each risk factor is in parentheses. E.g., the risk factor corresponding to whether the patient lives alone takes on the values 0 and 1; the risk factor corresponding to transferring takes on six values, from 0 through 5, with higher levels depicting greater dependence.

‡ All coefficients/odds ratios are significant at P<.10 using the likelihood ratio test for the hypothesis that the coefficient is zero. For this model, the five coefficients/odds ratios significant for .05<P<.10 are superscripted by "b," the one that is significant for .01<P<.05 is superscripted by "a," and the remainder are significant for P<.01. 90% CIs (confidence intervals) are given and odds ratios are considered significant at P<.10 (rather than the standard 95% CIs and P<.05 since the risk models were not used to assess impacts of risk factors on outcomes; rather, the model was used purely for predictive or risk adjustment purposes, resulting in the need to be inclusive).

§ The R^2 values are the squared correlations between predicted and observed values for all patients in the developmental and validation subsamples.

TABLE 3: Logistic Regression Model for the Outcome of Acute Care Hospitalization.

Risk Factor Measured at SOC/ROC*†	Coefficient‡	Odds Ratio‡	(90% CI)‡
Medicare as a payment source (0-1)	.07[a]	1.07	(1.02-1.13)
Inpatient discharge from hospital (0-1)	.30	1.35	(1.30-1.41)
Inpatient discharge from nursing home (0-1)	.20	1.22	(1.10-1.36)
Urinary catheter within past 2 weeks (0-1)	.24	1.27	(1.11-1.45)
Maximum severity rating among all diagnoses (0-4)§	.13	1.14	(1.11-1.18)
Number of diagnoses with severity rating ≥ 2 (0-6)§	.06	1.07	(1.05-1.08)
Moderate recovery prognosis (0-1)	-.16	.85	(0.80-0.91)
Good rehabilitation prognosis (0-1)	-.29	.75	(0.71-0.78)
Drug dependency at SOC (0-1)	.26[a]	1.30	(1.08-1.55)
Home sanitation hazard (clutter/soil/trash/vermin) (0-1)	.19	1.21	(1.13-1.29)
Patient lives alone (0-1)	.08	1.09	(1.04-1.14)
Patient has unpaid live-in help (0-1)	.08[a]	1.08	(1.02-1.15)
Vision impairment (0-2)	.05[a]	1.05	(1.01-1.09)
Pain interfering with activity (0-3)	.03	1.03	(1.01-1.05)
Stage of most problematic pressure ulcer (0-4)	.14	1.15	(1.10-1.21)
Stage 3-4 pressure ulcer(s) present (0-1)	.25[a]	1.28	(1.05-1.57)
Status of most problematic stasis ulcer (0-3)	.31	1.36	(1.30-1.42)
Surgical wound(s) present (0-1)	-.65	.52	(0.46-0.59)
Status of surgical wound (0-3)	.24	1.27	(1.19-1.36)
Dyspnea (0-4)	.18	1.19	(1.17-1.21)
Urinary catheter (0-1)	.30	1.34	(1.23-1.47)
Bowel ostomy (0-1)	.43	1.54	(1.37-1.74)
Anxiety Level (0-3)	.03[a]	1.03	(1.01-1.06)
Depression scale (0-5)	.09	1.09	(1.06-1.13)
Disability in grooming (0-3)	.08	1.08	(1.06-1.11)
Disability in ambulation (0-5)	.06	1.06	(1.03-1.08)
Disability in bathing (0-5)	.06	1.06	(1.05-1.08)
ADL assistance provided by caregiver (0-1)	-.14	.87	(0.84-0.91)
Prior (2 weeks ago) disability in laundry (0-2)	.13	1.14	(1.10-1.18)
Prior (2 weeks ago) disability in shopping (0-3)	.06	1.07	(1.04-1.10)
Prior (2 weeks ago) disability in transportation (0-2)	.22	1.25	(1.19-1.31)
Acute condition: mental/emotional (0-1)	.43	1.54	(1.38-1.72)
Acute condition: oxygen therapy (0-1)	.22	1.24	(1.17-1.32)
Acute condition: IV/infusion therapy (0-1)	.42	1.52	(1.38-1.66)
Acute condition: enteral/parenteral nutrition (0-1)	.64	1.89	(1.68-2.13)
Acute condition: cardiac/peripheral vascular (0-1)	.29	1.34	(1.28-1.41)
Acute condition: pulmonary (0-1)	.10	1.11	(1.05-1.16)
Acute condition: diabetes mellitus (0-1)	.10[a]	1.10	(1.03-1.78)
Acute condition: gastrointestinal disorder (0-1)	.17	1.19	(1.12-1.26)
Acute condition: contagious/communicable disease (0-1)	.28	1.32	(1.17-1.50)
Chronic condition: dependence in medication admin. (0-1)	.18	1.19	(1.14-1.25)
Chronic condition: chronic pain (0-1)	.20	1.23	(1.12-1.34)
Diagnosis: genitourinary system diseases (0-1)	.21	1.23	(1.17-1.30)
Diagnosis: skin/subcutaneous diseases (0-1)	.27	1.31	(1.22-1.41)
Diagnosis: neoplasms (0-1)	.49	1.64	(1.55-1.73)
Diagnosis: endocrine/nutritional/metabolic (0-1)	.25	1.29	(1.23-1.35)
Diagnosis: blood diseases (0-1)	.35	1.41	(1.33-1.50)
Diagnosis: circulatory system diseases (0-1)	.13	1.14	(1.09-1.20)

(continued on next page)

TABLE 3: Logistic Regression Model for the Outcome of Acute Care Hospitalization. (Cont'd)

Risk Factor Measured at SOC/ROC*†	Coefficient‡	Odds Ratio‡	(90% CI)‡
Length of stay: more than 31 days (0-1)	-.66	.52	(0.50-0.54)
Constant	-2.78		

Developmental sample R^2 = 0.165; Validation sample R^2s for 20 validation subsamples: Minimum = 0.145, Maximum = 0.165, and Mean = 0.155. ||

* SOC = Start of Care, ROC = Resumption of Care after inpatient stay. Risk factors pertain to SOC/ROC unless indicated otherwise.

† The number of values in the measurement scale for each risk factor is in parentheses.

‡ All coefficients/odds ratios are significant at P<.10 using the likelihood ratio test for the hypothesis that the coefficient is zero. For this model, the seven coefficients/odds ratios significant for .01<P<.05 are superscripted by "a," and the remainder are significant for P<.01. 90% CIs (confidence intervals) are given and odds ratios are considered significant at P<.10 (rather than the standard 95% CIs and P<.05 since the risk models were not used to assess impacts of risk factors on outcomes; rather, the model was used purely for predictive or risk adjustment purposes, resulting in the need to be inclusive).

§ A primary diagnosis and up to five other diagnoses are recorded for each patient at SOC/ROC. Each of the up to six diagnoses is rated according to the following five-level severity index:

0 - Asymptomatic, no treatment needed at this time
1 - Symptoms well controlled with current therapy
2 - Symptoms controlled with difficulty, affecting daily functioning; patient needs ongoing monitoring
3 - Symptoms poorly controlled, patient needs frequent adjustment in treatment and dose monitoring
4 - Symptoms poorly controlled, history of rehospitalizations

|| The R^2 values are the squared correlations between predicted and observed values for all patients in the developmental and validation subsamples.

The R^2 shown is the squared correlation between the predicted outcome probability[12] for each case and the actual outcome value (1 if the outcome occurred, 0 if the outcome did not occur). On the risk-adjusted outcome report, the "national" value shown for a given outcome is the mean of the patient-level predicted values for that outcome, while the "current" agency value is the actual proportion of patients who experienced the outcome.

5. Outcome Measures and Risk Adjustment: The Driving Force for OASIS

This section summarizes the approach to deriving the outcome measures that presently are used in the outcome reports for the national (CMS-sponsored) OBQI program. The discussion briefly highlights selected activities that took place in conducting the foundational research for OBQI as it was applied in the National and New York State Demonstration Trials.

Summary of Approach to Specifying Outcomes and Related Data Items: In keeping with Donabedian, a patient health status outcome is defined as a change in health status between two (or more) time points (Donabedian, 1980), typically admission and discharge for home health patients. Health status is broadly defined to include functional, physiologic, cognitive, and emotional/behavioral dimensions of health. An initial set of well over 500 patient health status outcomes was specified after reviewing the outcome and long-term care literature,

[12] The predicted outcome for patient j is calculated according to the formula $[1+\exp(a+\sum_i b_i r_{ij})]^{-1}$, where a is the logistic regression constant, b_i is the logistic regression coefficient for the i^{th} risk factor, and r_{ij} is the value of risk factor i for patient j.

conducting unstructured discussions on appropriate and useful patient outcomes with clinicians from all disciplines involved in home health care as well as representatives of regulatory and governmental programs, and surveying home health providers on outcomes. The draft set of outcomes was refined through several rounds of external review by a total of approximately 20 researchers in the long-term care field and 20 nationally recognized clinical experts from all disciplines involved in home health care. The multiple clinical and research panel reviews were designed to evaluate each outcome in terms of (1) validity as a quality indicator, (2) clinical utility for monitoring and improving care, (3) importance to patient health and well-being, and (4) expected measurement precision. Data items to measure the resulting outcomes were then developed using extant data sets, refining existing data items as needed, and developing new items when required items were nonexistent or inadequate. *It is important to note that the process of specifying, reviewing, and revising valid outcomes and outcome measures guided all efforts involved in developing what would eventually become the OASIS data set.*

Definitions of Key Terms Related to Quality and Quality Measures: The developmental work began with a set of literature-based, rigorous definitions of several terms related to quality.[13] Overall, the term "quality of care" was adopted as referring to a broad construct which in full generality is a pervasive attribute of health care that reflects its effectiveness relative to its objective(s) to assess, diagnose, prognosticate, cure, rehabilitate, stabilize, maintain, or palliate patient condition(s) and to ameliorate, prevent, minimize, or retard patient health problems. The presumption is made that each type of (home health) care has certain objectives. Quality of care refers to the extent to which these objectives are attained. When one speaks of quality of care, an implicit assumption is made that standards exist according to which the "goodness/adequacy" or "poorness/inadequacy" of care can be judged. Such standards can take the form of either expert opinion-derived norms,[14] or statistical norms reflecting the state of care provided at a given time (such as a prior period for a specific home health agency, or the current period when comparing a given agency with other agencies or a national sample). By definition, "quality of care" connotes a positive attribute of care, i.e., the more or higher the quality the better in terms of well-being of the patient.

Two other important terms used operationally in the context of the foundational research were "quality indicator" and "quality measure":

> As used in this research, the term quality indicator[15] also refers to a construct or an attribute of care that is conceptual in nature (not yet translated into a concrete attribute that is rigorously and precisely defined). A quality indicator refers to an attribute of care that can be used to gauge quality of care in a specific area. For example, improvement in patient wound healing, not necessarily specifying how one should actually measure the status of

---

[13] Numerous definitions exist of the terms given here, each of value in their own right in different contexts. Therefore, no pretense is made that the definitions of quality and quality-related constructs given here are either all-encompassing or among the most frequently used. Rather, they serve well the purposes of this research and are based on a thorough analysis of the terminology and clinical and theoretical principles that underpin the quality of home health care.

[14] In fact, although rarely used (despite the logic of doing so), the "experts" could be patients and their families or home caregivers. That is, particularly in the long-term, personal care field, such norms could be based predominantly if not exclusively on (realistic) patient or consumer preferences or expectations.

[15] Many do not distinguish between "quality indicator" and "quality measure" as the terms are used in this research. As discussed shortly, because of the appropriateness of first proceeding with the clinical and theoretical constructs of quality before and distinct from more specific quantitative and measurement issues, it was useful to distinguish clearly between quality indicators and quality measures.

the wound(s) or healing, is a quality indicator or construct that can reflect the quality of wound care. Thus, the term "quality of care" is a broad overarching construct, while the term "quality indicator" refers to a more specific construct that deals with a particular dimension of quality. The term "quality indicator" has been used in this research in a manner distinct from the term "quality measure" as discussed below.

A quality measure is in effect a rule (or the result of a rule) that assigns numeric values to a specific quality indicator. The essential distinction between quality indicators and quality measures is that quality measures take on numeric values, while quality indicators refer only to unquantified attributes of care related to quality. For example, improvement in ambulation is a quality indicator, while improvement in ambulation quantitatively reflected by a numeric change in a five-point ordinal mobility scale between admission and discharge is a quality measure. The time period over which an outcome measure is defined (e.g., from start of care until six weeks later, or from start of care until discharge) is the outcome interval or outcome episode. *One of the primary reasons the distinction between quality indicators and quality measures was used in this research is that operationally, certain types of clinicians and clinical panels are effective in developing and reviewing quality indicators, while other types of research and clinical panels are effective in developing and reviewing quality measures.* Therefore, a valid and useful quality measure takes on "values" (i.e., numbers), but is clinically and conceptually "rooted" in a quality indicator that is an unquantified attribute of care reflecting one of the many components of the overarching construct of quality (of care). Depending on how they are defined, quality measures and quality indicators can reflect either good/adequate/exemplary care or poor/inadequate/ inferior care.

In the foundational research that resulted in OBQI, process, structural, and outcome indicators and measures of quality were evaluated. The decision to focus on outcomes resulted from (1) the fact that health care is provided for the sake of influencing the well-being or health status of patients (i.e., the primary focus of health care is on patient outcomes) and (2) initial feasibility research had demonstrated that it was possible to measure patient outcomes of home health care. The terms "outcome" and "outcome indicator" were used synonymously in conducting the initial clinical and research panel reviews targeted on the content and substance of patient outcomes (i.e., not their measurement). Thereafter, the term "outcome measure" was used in order to transition from conceptual and clinical constructs that focus on patient well-being to the measurement of such constructs. The following definitions were used in this second stage of the foundational research:

An outcome (quality) measure is a quantification of a potential effect of care on the patient or those who informally support or care for the patient (since home care is sometimes directly targeted at such individuals). For example, (1) a dichotomous measure indicating whether a wound has healed between admission and six weeks after admission using a precisely defined scale to measure wound status, (2) a dichotomy indicating whether a patient was hospitalized potentially due to complications of care, (3) a quantification of whether a patient or home caregiver is satisfied with care received using a personal satisfaction scale, (4) a quantification of whether a home health patient or home caregiver is knowledgeable about certain aspects of self-care using a well designed pre/post approach to measuring knowledge of self-care, or (5) a dichotomy measuring whether pain interfering with activity has lessened between admission and discharge according to a pain scale developed for this purpose -- are outcome (quality) measures. For purposes of this research, outcome measures were subdivided into the three categories defined below.

An end-result or health status outcome measure reflects a change in patient condition or health status that is potentially due to the provision of care. In this regard, end-result

outcomes refer to changes/nonchanges in functional ability,[16] physiological conditions, emotional/mental health, cognitive health, behavioral health, or symptom distress that are intrinsic to the patient. For example, a quantification of change in ability to dress upper body between admission and three weeks after admission, a quantification of change between admission and discharge in dependence on intravenous medication (i.e., the physiological condition in this case is reflected by this dependency), and a quantification of symptom distress (e.g., pain present or absent) are end-result outcome measures.

An intermediate-result outcome measure reflects a nonphysiologic, nonfunctional, or, in general, nonhealth status outcome of care that is intrinsic to the patient, the patient's family/caregiver, or their behavior. For example, a quantification of the extent to which compliance occurs with respect to a medication regimen, a quantification of satisfaction with personal care services, or a dichotomy reflecting the extent of family/caregiver strain would be intermediate-result outcome measures.

Also referred to as surrogate end-result outcome measures in this research, utilization outcome measures are a quantification of health services use (or nonuse) that are potentially attributable to the (home health) care under consideration. Illustrations of utilization outcome measures include dichotomous measures of admission/nonadmission for urgent or emergent acute hospital care, length of hospital stay, and dichotomies reflecting discharge to the community or to a nursing home.

Although intermediate-result outcomes are desirable and valid from a conceptual perspective, the empirical testing that took place in the foundational research indicated that the data items necessary to measure intermediate-result outcomes were generally not of sufficient precision or reliability for operational purposes. As a result, the outcome measures that are presently used in OBQI are either end-result/health status outcome measures or utilization outcome measures.

A Description of the Outcome Measures Employed in the OBQI Demonstrations: Outcome measures and their associated data items were iteratively refined and reduced through field testing at over 200 home health agencies nationally during a period of several years in which (1) outcome measures and data items needed for the measures were continuously evaluated and revised clinically and empirically, (2) *new data items were added as a result of research to determine patient conditions and circumstances that would be essential for risk adjusting outcome measures for case mix differences among patient groups being compared*, (3) alternative outcome measurement approaches were examined for their practicality and utility in monitoring and managing impacts of care on patient well being, and (4) a continuous quality improvement (CQI) applications framework that would eventually be termed OBQI was developed and pilot tested in several home care agencies (Shaughnessy, Crisler, Schlenker, et al., 1994).

The final data set for measuring and risk adjusting outcomes was termed the Outcome and Assessment Information Set (OASIS) for home health care (HCFA, 2000; Humphrey, 2000).

---

[16] Although performance is traditionally used in measuring functioning rather than capacity or ability, all clinical panels that were convened agreed that for purposes of measuring outcomes of home care, it is imperative to (attempt to) measure ability. This is due to the nature of home care which often directly or indirectly focuses on attaining (or maintaining) the capacity of the patient to avoid institutionalization by acquiring (or retaining) the capacity to function at home. In fact, past research has shown that functional capacity and performance are extremely highly correlated. However, in home situations when the provider is aware that the patient is capable of independent functioning but is nonetheless receiving functional assistance (and therefore not directly demonstrating functional ability through current performance) from a temporary caregiver (e.g., daughter temporarily living with a patient to assist him/her), it still may be appropriate to discharge the patient to independent living.

This data set resulted exclusively from the need to validly and accurately measure and risk adjust patient outcome measures in home health care. It was designed to be integrated into a home health agency's start-of-care (SOC) and follow-up patient assessment forms, replacing like items with OASIS items. Despite the fact that the more precisely worded OASIS data items required more space in assessment forms, agencies already performing a clinically appropriate comprehensive assessment of each patient added little, if any, new content to their SOC assessment form when they replaced extant items with OASIS items. A subsequent (double-blind) time survey showed once clinical staff became familiar with OASIS items, average time required for patient assessments returned to that required before OASIS -- for clinicians that had previously been conducting comprehensive assessments as required under Medicare. The time study is discussed in more detail elsewhere (CHSR, Volume 3, Supporting Document 2, 2002).

Consistent with Medicare certification requirements, OASIS items were completed at SOC for each adult (18 years or older), nonmaternity patient. OASIS data were collected approximately every 60 days until and including time of discharge. A 60-day interval, which resulted from testing data collection intervals of varying lengths (Shaughnessy, Crisler, Schlenker, et al., 1994), was optimal for the combined objectives of provider efficiency and clinical and statistical utility of outcome measurement. This length of time allowed the data items to be used effectively for assessing patient home care needs, measuring outcomes as changes in health status between SOC and follow-up points, and risk adjusting outcome measures to take into consideration case mix differences between an agency's patients and its comparison group.

The foundational research activities based on (considerable) clinical input, empirical testing, and expert review converged, for the sake of parsimony and clinical utility, on the 41 dichotomous outcome measures enumerated in Table 4. These were the measures used in the National and New York State OBQI Demonstration Trials. In fact, other outcome measures were developed and can be computed from the OASIS data set. For example, it is possible to use composite outcome measures that span or summarize the information in several of the individual outcome measures in Table 4. In this regard, it is possible to use a few composite functional outcome measures that capture the essence of the individual functional measures (combined) in Table 4. The reason for using the Table 4 outcome measures for OBQI (as opposed to more general reporting or evaluation purposes) is discussed shortly. Every OASIS data item other than patient identifiers is used either to measure or risk adjust at least one, but typically many, of the 41 outcomes. The majority of the outcome measures focus on patient functioning because of the emphasis in home care on assisting patients to become or remain sufficiently independent to stay in their home environment, thereby avoiding institutional long-term or acute care. The more technical or specific skilled services provided by home health agency staff such as wound care, IV administration, infection control, cardiac evaluation and monitoring, and cognitive assessment often are directed toward or culminate in assisting patients to (resume) function(ing) as well as possible at home.

Each health status outcome measure is either an improvement or stabilization measure as defined in the first footnote of Table 4. At times, the intent of home health care is to prevent decline or slow the rate of decline in health status. Hence, in addition to improvement measures, Table 4 contains selected stabilization measures that are useful to evaluate outcomes. The research considerations and principles that shaped the approaches to outcome measurement in this program are in discussed an earlier article (Shaughnessy, Crisler, Schlenker, 1994).

Composite vs. Focused Outcome Measures and OASIS Reliability: It is possible to employ composite or aggregate outcome measures corresponding to multiple domains of health such

TABLE 4: Outcome Measures Used in OBQI Demonstrations.

Health Status Outcome Measures*

Functional: Activities of Daily Living

Improved in:
- Ambulation/Locomotion
- Dressing Upper Body
- Dressing Lower Body
- Grooming
- Bathing
- Eating
- Toileting
- Transferring

Stabilized in:
- Grooming
- Bathing
- Transferring

Functional: Instrumental Activities of Daily Living

Improved in:
- Management of Oral Medications
- Light Meal Preparation
- Laundry
- Housekeeping
- Shopping
- Telephone Use

Stabilized in:
- Management of Oral Medications
- Light Meal Preparation
- Laundry
- Housekeeping
- Shopping
- Telephone Use

Physiologic

Improved in:
- Pain Interfering with Activity
- Number of Surgical Wounds
- Status of Surgical Wounds
- Dyspnea
- Urinary Tract Infection
- Urinary Incontinence
- Bowel Incontinence
- Speech or Language

Stabilized in:
- Speech or Language

Emotional/Behavioral

Improved in:
- Anxiety Level
- Behavioral Problem Frequency

Stabilized in:
- Anxiety Level

Cognitive

Improved in:
- Confusion Frequency
- Cognitive Functioning

Stabilized in:
- Cognitive Functioning

Utilization Outcome Measures†
- Acute Care Hospitalization
- Discharge to Community
- Emergent Care

* A patient improves when the scale value for the health attribute under consideration shows an improvement from one time point to the next (i.e., between admission and discharge). If the patient is less disabled or dependent at discharge than at start of care, then the patient has improved and the improvement outcome measure takes on the value "1;" otherwise, it takes on the value "0." If the patient is at the most independent or "healthiest" extreme of the scale, it is impossible to improve, and therefore the measure is not defined for such patients.

A patient stabilizes when the scale value for the health attribute under consideration shows nonworsening in patient condition. If the patient is no more disabled or dependent (that is, has not worsened) at discharge than at start of care, then the patient has stabilized and the stabilization outcome measure takes on the value "1," whereas if the patient worsened, the stabilization measure takes on the value "0." If the patient is not able to worsen according to the scale (i.e., is at the most dependent or "sickest" extreme of the scale), then the measure is not computed for this particular patient. Exclusions result in differing sample sizes for each health status outcome.

As an illustration, the ambulation/locomotion scale in OASIS is specified as follows:

Ability to SAFELY walk, once in a standing position, or use a wheelchair, once in a seated position, on a variety of surfaces.

0 - Able to independently walk on even and uneven surfaces and climb stairs with or without railings (i.e., needs no human assistance or assistive device).
1 - Requires use of a device (e.g., cane, walker) to walk alone or requires human supervision or assistance to negotiate stairs or steps or uneven surfaces.
2 - Able to walk only with the supervision or assistance of another person at all times.
3 - Chairfast, unable to ambulate but is able to wheel self independently.
4 - Chairfast, unable to ambulate and is unable to wheel self.
5 - Bedfast, unable to ambulate or be up in a chair.

If the patient is classified at level "0" at start of care, the improvement measure for ambulation/locomotion is not computed. If the patient is classified at level "5," the stabilization measure is not computed.

† Utilization outcomes take on the value "1" if the indicated event occurs; otherwise, they take on the value "0." Unlike the aforementioned improvement and stabilization outcomes, these outcomes are defined for all patients. For example, if the patient is hospitalized (as an inpatient, for a period of at least 48 hours), the hospitalization measure takes on the value "1;" otherwise, it takes on the value "0."

as the aforementioned global measure(s) encompassing several or all domains of functioning simultaneously. Such outcome measures can use near-continuous or at least polychotomous (ordinal) scales. Dichotomous outcomes corresponding to individual domains of health status were chosen for OBQI applications for two reasons. First, they are more readily understood by home care clinicians. Second, they are more straightforward to use in enhancing patient outcomes than more complex composite outcome measures. Use of the composite measures necessitates determining which specific outcome domains or health status attributes must be further examined to improve the composite measure. This naturally takes the clinician back to using results from dichotomous measures corresponding to specific health status attributes. As mentioned, the utilization outcomes in Table 4 can be regarded as surrogate health status outcomes. While the 41 outcome measures in this table have been selected for OBQI applications at the present time, research to improve upon these outcomes for OBQI and other applications is ongoing. For example, composite or aggregate outcomes are being analyzed for purposes of more general evaluation and informational purposes (i.e., beyond OBQI).

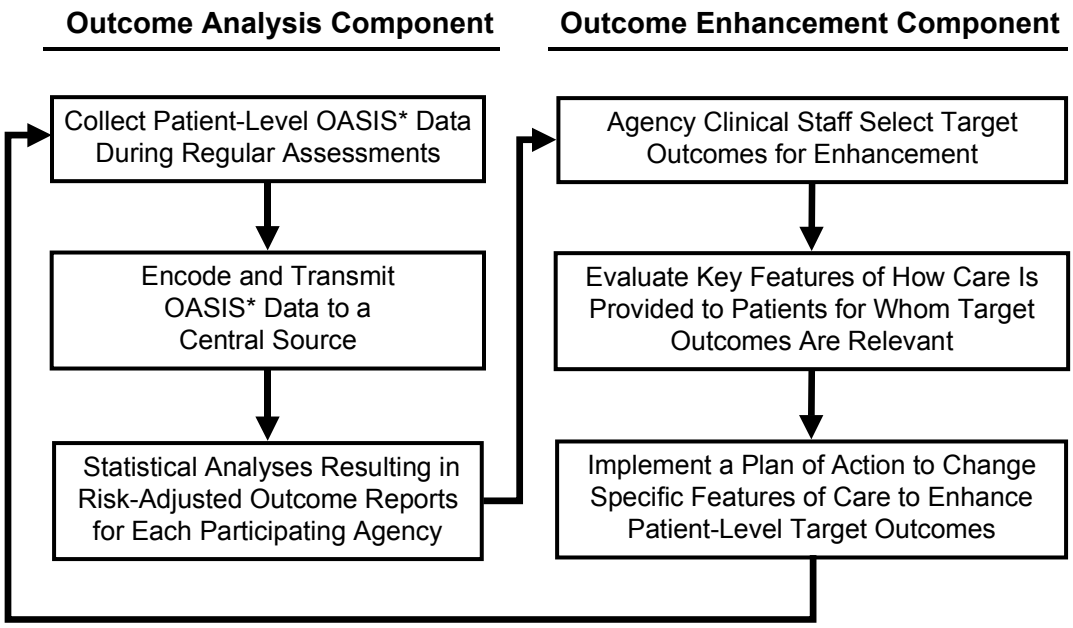
Information on reliability of OASIS data items is available in CHSR, Volume 3, Supporting Document 2 (2002) and elsewhere (Shaughnessy, Crisler, Schlenker, et al., 1997; Madigan and Fortinsky, 2000). An outcome measure was judged to be reliable for OBQI if the underlying data item used to compute the outcome had an interrater reliability coefficient (weighted kappa) greater than 0.60. (The raters included two skilled nurses trained in using OASIS who did duplicate assessments within 24 hours of one another, while randomizing which rater was first, for 66 patients from five agencies.) Only two of the items used to compute the 38 health status outcome measures in the OBQI outcome reports had coefficients less than 0.60 (both were 0.54). Twenty-five outcome measures had reliability coefficients greater than 0.70. OASIS reliability testing and refinements to improve reliability will continue as CMS and others use OASIS and OBQI nationally. The current version of OASIS, with an explanatory prologue and instructions for using OASIS, is provided in CHSR, Volume 3, Supporting Document 4 (2002).

6. Summary Explanation of OBQI and Risk-Adjusted Outcome Comparisons

The essential features of the CQI applications framework employed in the OBQI demonstration trials were developed and refined in a pilot project involving several home health agencies in the early 1990s. A detailed explanation of OBQI can be found in CHSR, Volume 3, Supporting Document 5 (2002). This framework includes the outcome analysis and outcome enhancement components depicted in Figure 1. The outcome analysis component begins with collecting, electronically encoding, and transmitting OASIS data to a central source (the appropriate state agency in the case of the current national program; OASIS data then are transmitted to the national OASIS repository from the individual states). In the demonstrations, outcome, case mix, and adverse event reports were returned to each agency on an annual basis. In the current national program, reports are available to agencies on request. Outcome reports may be selected to cover a 12-month time period only, but adverse event reports may be selected for as little as one month. Most important is the All Patients' Outcome Report that permits agency staff to analyze their patient outcomes aggregated to the agency level. Two additional outcome reports for orthopedic and cardiac patients were provided in the demonstrations when agency-specific sample sizes were adequate. The outcome reports provide a comparison of agency performance relative to (1) a national reference or benchmark population and (2) that agency's prior year outcomes.

As discussed, the 41 outcome measures presented in Table 4 were selected from a larger set of outcomes for the all-patient reports. The case mix report contains an overview of the agency's admitting case mix on about 140 case mix factors for the current time period or year relative to the agency's prior time period and the (national) reference or benchmark group. The

FIGURE 1: OBQI Applications Framework.

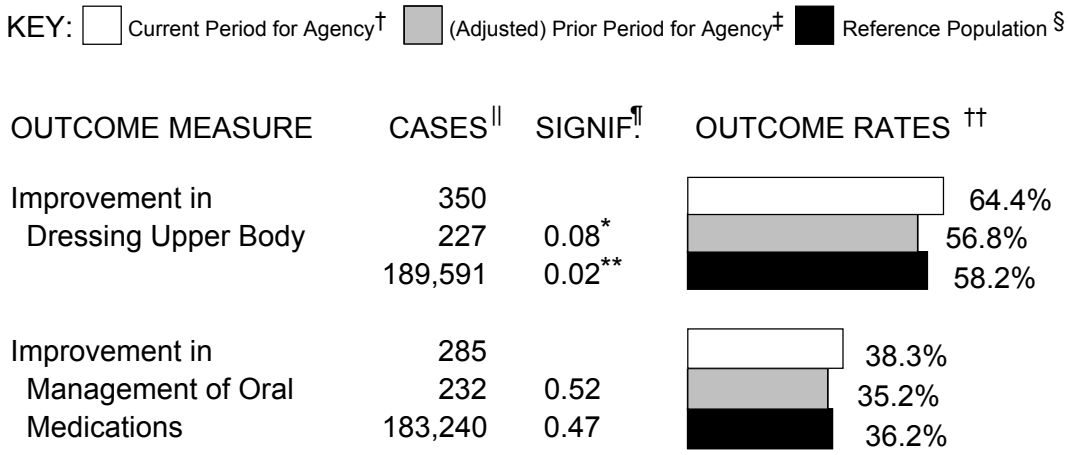


adverse event report provides a profile on low-frequency adverse patient outcomes such as emergent care for hypo/hyperglycemia. Incidence rates for such events are too low for meaningful risk adjustment. Their occurrence generally requires review and scrutiny on a case-by-case basis. Figure 2 provides an illustration from an outcome report. The footnotes in Figure 2 explain the terms and computations involved in the report. A complete copy of the outcome report series for a hypothetical agency is presented in CHSR, Volume 3, Supporting Document 6 (2002). It contains sample outcome, case mix, and adverse event reports as well as a patient satisfaction report distributed during the demonstrations.

In the second component of the OBQI applications framework, outcome enhancement, home health agencies are given considerable latitude to conduct their own CQI activities. These include process-of-care investigations that culminate in the development and implementation of plans of action specifying how care behaviors will be changed to enhance participating outcomes. In the demonstration trials, upon conclusion of the first year, each participating agency received an outcome report showing its performance during that year relative to all demonstration agencies (no prior period comparison was possible the first year). After reviewing the report, various methods were implemented by agency staff to assess and change specific care behaviors related to target outcomes of their choosing. Thus, agencies produced written plans of action for their target outcomes at the beginning of the second demonstration year. These plans documented the target outcome, care behaviors being targeted, how they would be changed, who would be responsible for implementing the changes, and how the changes would be monitored on a continuing basis. The outcome enhancement process was repeated at the start of each succeeding demonstration year when the annual outcome report was received.

Accurate and uniform data on patient health status are critically important to OBQI. In addition, accuracy and uniformity of patient information introduce far more rigor into the

FIGURE 2: Excerpt (Using 2 of 41 Outcomes) from an Agency-Level Risk-Adjusted Outcome Report.



† The unshaded or "current" rates pertain to the actual or observed improvement rates for each outcome during the most recent period (usually about 12 months).

‡ The partially shaded or "(adjusted) prior period" rates are the risk adjusted comparative rates for the period preceding the most recent one for the given agency and are computed by

$$\hat{y}(\text{adj prior}) = y_p(\text{obsv}) + [\hat{y}_c(\text{pred}) - \hat{y}_p(\text{pred})],$$

where $y_p(\text{obsv})$ = the observed outcome rate for the prior period for the agency; $\hat{y}_c(\text{pred})$ = the predicted rate for the current period obtained by substituting all patients (episodes) eligible for the outcome for the agency into the outcome's logistic regression model based on pooled data from the prior and current period, and computing the average or rate across all such predicted values; and $\hat{y}_p(\text{pred})$ = the predicted rate for the prior period obtained by substituting all patients eligible for the outcome for the agency into the outcome's logistic regression model based on pooled data from the prior and current period -- and computing the average or rate across all such predicted values. Intuitively, the expression $[\hat{y}_c(\text{pred}) - \hat{y}_p(\text{pred})]$ represents the outcome rate adjustment to the baseline (observed prior) rate that is due to the difference or change in risk factors between the current and prior periods.

§ The darkened or "reference" rates are the risk-adjusted comparative reference sample rates for each outcome computed by calculating the mean of the predicted or expected values for each nonexcluded patient admitted to the agency under consideration. These are obtained by substituting each patient's values for the risk factors into the logistic regression model for that outcome (which was estimated using all patients in the reference sample), and averaging the predicted values across all the agency's (nonexcluded) patients for the outcome.

|| The "cases" column contains the number of outcome episodes for the time period or the reference group to which the bar to the far right of the case count corresponds (in the instance of the reference group, this is the number of cases on which the logistic regression model for the outcome was estimated). The number of cases for each of the three samples reflects the number of patients not excluded from the outcome measure computation for the indicated period or group, using the exclusion criteria given in the first footnote of Table 4.

¶ The "signif." column contains the statistical significance level of the test of the null hypothesis that the outcome rate for the current period is the same as the comparative rate (for either the adjusted prior period or the reference group). For the current vs. adjusted prior comparison, this is the significance level for testing the hypothesis that two sample binomial proportions are the same. In the case of small samples (N<30), it is the significance level for Fisher's exact test, and for larger samples, it is the significance level for the chi-square approximation to this test. In comparing the current vs. reference population rate, it is the significance level for the one-sample binomial test that a sample proportion is equal to a population proportion. Several statistical tests were examined theoretically and experimentally before determining that these two binomial tests are the most informative and sensitive for generating these significance levels. Single and double asterisks are used on the outcome report to convey significance for P<.10 and P<.05, respectively, so that the report can be scanned to determine where there might be significant differences (either favorable or unfavorable) in outcomes.

†† The outcome rates for the two time periods and the reference groups are depicted by the length of the respective bars for the outcome. Their numeric values are given to the right of each bar.

assessment process than had been customary in home health care. Further, the OBQI approach entails a considerably different type of thinking and philosophy than is typically used by home care providers. Such thinking requires analysis that links care to specific outcomes in managing care behaviors and coordinating changes in care behaviors that involve multiple providers. Separate training programs accompanying the demonstration trials focused on incorporating OASIS into comprehensive assessments and using OBQI to evaluate and enhance patient outcomes. The national training programs that CMS has sponsored and is continuing to sponsor are modeled after these earlier programs.

7. Future Plans: Continued Refinement of Risk Adjustment and Outcome Measures

The risk adjustment and outcome measurement methodologies described in this document are being refined and will evolve in areas such as those indicated below.

a. Risk adjustment models are being reestimated at least yearly to ensure that they continue to be current as national risk-adjusted outcome reports are produced for each home health agency nationally. Because of data constraints and methodological issues, selected outcomes were not risk adjusted in the 2002 national outcome reporting cycle. As these constraints and issues are addressed, such outcomes will be either risk adjusted or modified so risk adjustment is possible.

b. The availability of OASIS data for a large number (millions) of cases makes it possible to experiment with more extensive (in terms of numbers of risk factors) and complex (in terms of functional form) logistic regression models. This experimentation will result in (possibly substantial) improvements in predictive accuracy and/or enhanced clinical validity and utility.

c. New risk factors will be studied, including transformations of (combined) current risk factors, agency-level characteristics, and market-level/geographic-area characteristics. Some higher level characteristics (e.g., agency-level variables) may be used as actual risk factors for outcome reporting. Others will not be used as risk factors. Instead they will be studied to explain variations in patient outcomes as potential causative agents that can be influenced by provider initiatives and policy changes, or at least taken into consideration in the context of such initiatives and changes.

d. Alternative risk adjustment methodologies will be studied and tested using OASIS data with the goal of developing more effective risk adjustment approaches that also are intuitively understandable to clinicians and other users including consumers. These will encompass but not be limited to:

   (1) Propensity scoring to render test-control comparisons more precise or appropriate (Normand, Landrum, Guadagnoli, et al., 2001; Rubin and Thomas, 1996),

   (2) Extended simultaneous uses of multiple goodness-of-fit statistics such as R^2s, C-statistics, and deciles of risk,

   (3) Further estimation of Markovian transition models to produce overall risk-adjusted approaches for a single outcome by modeling transitions from level to level within a given outcome,

   (4) Risk adjustment using different units of analysis, including agency-level and patient-level analyses possibly combined through applying hierarchical analyses, and using different

types of statistical estimation and regression methodologies (Raudenbush and Byrk, 2002; Normand, Glickman, and Gatsonis, 1997),

(5) More extensive testing of the general method of data handling (GMDH) (Ivakhnenko, 1991) and simulated neural networks (SNNs) (Tu, 1996; Grigsby, Kooken, and Hershberger, 1994), and

(6) Extended "propensity" analysis focused on k-to-1 patient-level matching using distance minimization (between test and comparison sample risk factor centroids and covariance/correlation matrices) that extend current methods (Rosenbaum and Rubin, 1985).[17]

e. Alternative outcome measures and risk factors will be examined as the OASIS data set is revised through continued research to improve the reliability and validity of items that are currently part of OASIS (and to develop new or substitute items as needed).

f. New outcome measures, including composite measures that span multiple health domains, will be investigated for assessing or evaluating agency performance from various non-OBQI perspectives such as payer, multiprovider, regulatory, and consumer perspectives. [As noted, except for a few utilization outcomes such as hospitalization, health problem- or attribute-specific measures have been emphasized to date because they are clinically more meaningful for the OBQI methodology.]

Risk adjustment is a critically important component of outcome reporting. Without adequate data items to measure potential risk factors, the ability of CMS to continue to evaluate and improve risk adjustment would be seriously impaired. The OBQI and OASIS Change and Evolution Program (OCEP) recently implemented by CMS bears considerable promise in this regard. As experience is gained with OASIS, it will be possible to make occasional changes in the data set (modifying, deleting, or supplementing data items) that will be in the best interest of continually improving not only outcome measures, but equally important, the capacity to risk adjust such measures for OBQI and public reporting applications.

[17] The six indicated areas (in a-f) are illustrative of several of a variety of methods and issues being addressed.

REFERENCES

Agresti A. *Categorical Data Analysis*. New York, NY: John Wiley and Sons; 1996.

CHSR, Volume 3, Supporting Documents, 2002 [several supporting documents from Volume 3 of the four-volume report series written by the faculty and staff of the Center for Health Services Research (CHSR) are cited in the text. See the below reference by Shaughnessy, Crisler, Hittle, et al., for the complete citation to the four-volume set].

Donabedian A. Explorations in Quality Assessment and Monitoring. Volume I: The Definition of Quality and Approaches to Its Assessment. Ann Arbor, MI: Health Administration Press;1980.

Flack VF, Chang PC. Frequency of selecting noise variables in subset regression analysis: A simulation study. *American Statistician.* 1987;41:84-86.

Freedman DA. A note on screening regression equations. *American Statistician.* 1983;37:152-155.

Grigsby J, Kooken R, Hershberger J. Simulated neural networks to predict outcomes, costs, and length of stay among orthopedic rehabilitation patients. *Archives of Physical Medicine and Rehabilitation.* 1994;75:1077-1081.

Harrel FE, Lee KI, Mark DB. Tutorial in biostatics: Multivariable prognostic models: Issues in developing models, evaluating assumptions and measuring and reducing errors. *Statistics in Medicine.* 1996;15:361-387.

Health Care Financing Administration, US Department of Health and Human Services (2000). OASIS data sets: OASIS–B1 (8/2000) final version. Available at http://www.hcfa.gov/medicaid/oasis/oasisdat.htm.

Hosmer DW, Lemeshow S. *Applied Logistic Regression.* 2nd ed. New York, NY: John Wiley and Sons; 2000.

Humphrey CJ, ed. Special focus issue: OASIS and OBQI. *Home Healthcare Nurse.* 2000;18(4):217-280.

Iezzoni LI. *Risk Adjustment for Measuring Health Outcomes*. Ann Arbor, MI: Health Administration Press; 1994.

Ivakhnenko AG. An inductive sorting method for the forecasting of multidimensional random processes and events with the help of analogues forecast complexing. *Pattern Recognition and Image Analysis.* 1991;1:99-108.

Lee K, Koval JJ. Determination of the best significance level in forward stepwise logistic regression. *Communication in Statistics B.* 1997;26:559-575.

Madigan EA, Fortinsky RH. Additional psychometric evaluation of the outcomes and assessment information set (OASIS). *Home Health Care Services Quarterly.* 2000;18(4):49-63.

Mickey J, Greenland S. A study of the impact of confounder-selection criteria on effect estimation. *American Journal of Epidemiology.* 1989;129:125-137.

Mittlböck M, Schemper M. Explained variation for logistic regression. *Statistics in Medicine.* 1996;15:1987-1997.

Normand S-LT, Glickman ME, Gatsonis CA. Statistical methods for profiling providers of medical care: Issues and applications. *Journal of the American Statistical Association.* 1997;92:803-814.

Normand S-LT, Landrum MB, Guadagnoli E, Ayanian JZ, Ryan TJ, Cleary PD, McNeil BJ. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: A matched analysis using propensity scores. *Journal of Clinical Epidemiology.* 2001;54(4):387-398.

Osius G, Rojek D. Normal goodness-of-fit tests for multinomial models with large degrees-of-freedom. *Journal of the American Statistical Association.* 1992;87:1145-1152.

Raudenbush SW, Byrk AS. *Hierarchical Linear Models.* 2nd ed. Thousand Oaks, CA: Sage Publications, Inc.; 2002.

Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician.* 1985;39:33-38.

Rothman KJ, Greenland S. *Modern Epidemiology.* 3rd ed. Philadelphia, PA: Lippincott Raven; 1998.

Rubin DB, Thomas N. Matching using estimated propensity scores: Relating theory to practice. *Biometrics.* 1996;52:249-264.

Shaughnessy PW, Crisler KS, Hittle DF, et al. Four-volume report series entitled *OASIS and Outcome-Based Quality Improvement in Home Health Care: Research and Demonstration Findings, Policy Implications, and Considerations for Future Change*, Denver, CO: Center for Health Services Research, University of Colorado Health Sciences Center, February 2002.

Volume 1: Policy and Program Review
Volume 2: Research and Technical Overview
Volume 3: Research and Clinical Supporting Documentation
Volume 4: OASIS Chronicle and Recommendations

Shaughnessy PW, Crisler KS, Schlenker RE, et al. Outcomes across the care continuum. Home health care. *Medical Care.* 1997;35(11 Suppl):NS115-NS123 and 35(12):1225-1226.

Shaughnessy PW, Crisler KS, Schlenker RE, et al. Measuring and assuring the quality of home care. *Health Care Financing Review.* 1994;16(1):187-210.

Shaughnessy PW, Hittle DF, Crisler KS, et al. Improving patient outcomes of home health care: Findings from two demonstration trials of outcome-based quality improvement, *Journal of the American Geriatrics Society.* 2002; 50:1354-1364.

Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology.* 1996;49:1225-1231.