



**National Quality Forum  
'Hospital Performance: Additional Priority Areas'**

**CareScience Risk Adjusted Average Length of  
Inpatient Hospital Stay Measure**

**Name of Organization:** Premier, Inc.  
**Name of Contact:** John Martin  
**Address:** 3600 Market Street, 7<sup>th</sup> Floor  
Philadelphia, PA 19104  
**Telephone:** (215) 689-2240  
**Fax:** (215) 387-9406  
**Email Address:** [John.Martin@Premierinc.com](mailto:John.Martin@Premierinc.com)

## Table of Contents

<b>I. CARESCIENCE RISK ADJUSTED AVERAGE LENGTH OF STAY (ALOS) MEASURE DEFINITION (NUMERATOR, DENOMINATOR, AND DATA ELEMENT DEFINITION)</b> .....	<b>3</b>
<b>II. DISEASE GROUP INCLUSIONS AND EXCLUSIONS</b> .....	<b>4</b>
<b>III. DATA SOURCE</b> .....	<b>4</b>
<b>IV. LENGTH OF STAY MEASURE ALLOWABLE VALUES</b> .....	<b>4</b>
<b>V. DATA ANALYSIS LOGIC AND METHOD</b> .....	<b>5</b>
<b>VI. RISK-ADJUSTMENT METHOD AND ASSOCIATED DATA ELEMENTS</b> .....	<b>6</b>
RISK MODEL INDEPENDENT VARIABLES (RISK FACTORS).....	6
FUNCTIONAL FORM AND RISK MODEL STRATIFICATION.....	7
SAMPLING FOR THE RISK MODEL CALIBRATION DATASET.....	7
RISK MODEL CALIBRATION.....	7
DATA SOURCES – PUBLIC AND PRIVATE DATA.....	8
<b>VII. COHORT DEFINITION AND SAMPLING METHOD</b> .....	<b>9</b>
COHORT DEFINITION.....	9
SAMPLING METHOD.....	9
<b>VIII. ALOS MEASURE REPORTING FORMAT</b> .....	<b>9</b>
INTERPRETING RESULTS: STATISTICAL SIGNIFICANCE.....	9
AGGREGATION.....	10
<b>IX. CARESCIENCE LOS MEASURE PURPOSE AND DURATION OF USE</b> .....	<b>10</b>
<b>X. CARESCIENCE LOS MEASURE AUDIT STUDIES</b> .....	<b>11</b>

# I. CareScience Risk Adjusted Average Length of Stay (ALOS) Measure Definition (Numerator, Denominator, and Data Element Definition)

The CareScience ALOS measure represents the number of excess in-hospital days in a given inpatient population,  $j$ . The results represent the difference between the number of expected hospital days subtracted from the number of observed hospital days.

Length of Stay is defined by Premier as the number of full days a patient stays in the hospital. The patient population can be aggregated as any grouping of patients (e.g. by hospital, by physician, by diagnosis code, by procedure, by DRG, etc.) It is calculated as the difference between discharge date and admission date.

Certain outcome measures, notably costs per case and length-of-stay (LOS), are distributed with a rightward (positive) skew. Applying linear regression to models with skewed dependent variables gives rise to a number of pathologies, including inefficient, often biased, parameter estimates and predictions outside logical bounds, such as negative values for LOS and costs. When outcome measures are not symmetrically distributed, analysis of performance can be disproportionately influenced by outliers and special or extreme cases. This phenomenon can require a manual procedure for identifying and removing outliers, a subjective technique at best.

A more robust solution is to take the natural log of the dependent variable, which results in an approximately symmetric distribution and contracts the outliers inward toward the center of the data. It also ensures that all predicted values will be positive. Therefore, in modeling the expected LOS Premier uses the natural log of the patient LOS. The actual model works with the geometric mean (GM) of LOS and is expressed as:

$$\text{CareScience LOS measure} = \exp(\bar{y}_j) - \exp(\hat{\bar{y}}_j)$$

Where:

Observed GM mean is  $\exp(\bar{y}_j)$  and  $\bar{y}_j = \frac{1}{n_j} \sum_{ik}^{n_j} y_{ijk}$

Expected GM mean is  $\exp(\hat{\bar{y}}_j)$  and  $\hat{\bar{y}}_j = \frac{1}{n_j} \sum_{ik}^{n_j} \hat{y}_{ijk}$

$$y_{ijk} = \ln(\text{LOS}_{ijk})$$

$$\hat{y}_{ijk} = \ln(\text{LOS}_{ijk} \text{ expected})$$

$i$  = patient

$j$  = provider or grouping

$k$  = icd9 diagnosis (3 digit)

## II. Disease Group Inclusions and Exclusions

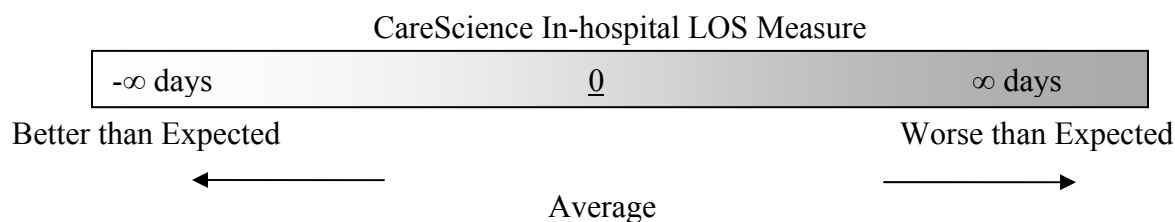
The CareScience LOS measure methodology is flexible enough to accommodate any range of diagnosis group definitions involving different principal and secondary diagnosis codes. The shortest valid LOS is one day. If a patient is admitted and discharged on the same day, LOS is counted as one day. The top 1% within each disease strata will be trimmed to remove outliers. Rather than excluding patients based on narrow criteria, our model takes into account various factors such as discharge disposition to risk adjust the LOS. To that end, we account for areas of affecting the LOS by including risk factors and their interactions, rather than using exclusion criteria, to ensure the results stemming from this measure are more robust and represent the true patient population with the specified disease.

## III. Data Source

The CareScience LOS measure's methodology uses data supplied by hospitals and can be easily applied to UB-92/UB-04 type discharge data. The methodology described and its corresponding data types have been tested, validated, commercialized, and published in peer-reviewed journals<sup>1,2,3</sup>. The methodology is owned by the University of Pennsylvania and is licensed exclusively to Premier, however, it is publicly available.

## IV. Length of Stay Measure Allowable Values

The CareScience LOS measure returns a continuum of values based on the deviation from the expected LOS. Values less than 0 indicate a LOS that is lower than expected given patient severity. Conversely, LOS measure values greater than 0 indicate a LOS that is higher than expected given patient severity.



<sup>1</sup> Brailer, David J.; Kroch, Eugene A.; Pauly, Mark V.; Huang, Jianping. *Medical Care*. (1996) Vol. 34, No. 5, 490-505.

<sup>2</sup> Azimuddin, Khawaja; Rosen, Lester; Reed, James F. *Diseases of the Colon and Rectum*, April 2001.

<sup>3</sup> Pauly, Mark V.; Brailer, David J.; Kroch, Gene; Even-Shoshan, Orit. *American Journal of Medical Quality*. Fall 1996

## V. Data Analysis Logic and Method

Evaluation of LOS as an indicator of hospital (or physician) service quality requires comparison of the hospital's (or physician's) observed in-hospital LOS to their expected in-hospital LOS as suggested by their patient case mix. The expected LOS for any facility or grouping of patients is based on the characteristics of those patients as they reflect patient severity. A model of the relationship between these patient characteristics and LOS is estimated statistically through a regression analysis of a sample from a defined population of hospital discharges.

The Premier approach to analyzing LOS is to calibrate our patient risk model on a nationally-based sample of discharges.<sup>4</sup> The aim is to construct a set of parameter estimates (i.e., *beta* coefficients) that can be used to predict LOS for any set of patients. Predicted LOS can be compared to the actual rates to evaluate performance for any set of patients or for a particular facility, service line, diagnosis, age grouping, or treating physician. We therefore label these parameter estimates "universal" *beta* coefficients.

This approach has a number of advantages over the alternative method of calibrating the risk-adjustment on an individual hospital or hospital system. First, outcome risks that are generated by a set of universal *beta* coefficients can be compared across all facilities and all physicians wherever they practice. Second, any set of discharges can be analyzed, however small, since the model parameters do not need to be estimated from the analysis data set. Third, having a set of universal *beta* coefficients allows real-time processing of discharges.

Each patient is assigned an in-hospital LOS risk (expected) score. Patients can therefore be analyzed by any grouping (e.g. by diagnosis, by disease group, by provider, by DRG, etc.).

---

<sup>4</sup> Pauly MV, Brailer DJ, and Kroch EA, "The Corporate Hospital Rating Project: Measuring Hospital Outcomes from a Buyers Perspective," *American Journal of Medical Quality*, **11**(3):112-122.

## VI. Risk-Adjustment Method and Associated Data Elements

The purpose of the CareScience risk model is to generate expected LOS under typical care based on a patient's health status and other relevant patient characteristics. Patient-level LOS risks are assessed via a stratified disease-specific logistic regression model. The strata are roughly based on 3-digit level ICD-9-CM diagnosis codes. LOS risk scores and their standard errors can be used to assess actual provider experience by reporting how observed LOS differs from their expected LOS or risk.

### **Risk Model Independent Variables (Risk Factors)**

The following patient characteristics and socioeconomic factors comprise the set of regressors or classes of independent variables for the CareScience risk model.

**Age** (*quadratic form*)

**Birth weight** (*quadratic form*)

**Sex** (*female, male, unknown*)

**Race** (*white, black, other, unknown*)

**Hispanic flag**

**Income** (*median household income within a zip code reported by US Census Bureau*)

**Distance traveled** (*the centroid-to-centroid distance between the zip code of the household and the zip code of the hospital or provider, represented as a relative term*)

**Principal diagnosis** (*terminal or three digit ICD-9-CM code, where statistically significant*)

**CACR<sup>5</sup>, comorbidity scores** (*count of comorbidities within each of five severity categories on the CACR Likert scale*)

**Defining diagnosis** (*three digit ICD9-CM code for neonates*)

**Cancer status** (*benign, malignant, carcinoma in situ, history of cancer, derived from secondary diagnoses*)

**Chronic disease and disease history** (*terminal digit ICD9-CM diagnosis codes, such as diabetes, renal failure, hypertension, chronic GI, chronic CP, obesity, and history of substance abuse*)

**Valid procedure** (*terminal ICD9-CM procedure codes, where clinically relevant and statistically significant*)

**Time trend factor for cost and charge outcomes** (*to control for inflation specific to each disease in the inpatient hospital setting, derived from discharge date*)

**Admission source** (*Physician Referral, Clinic Referral, HMO Referral, Transfer from a Hospital, Skilled Nursing Facility or Another Health Care Facility, Emergency Room, Court/Law Enforcement, Newborn - Normal Delivery, Premature Delivery, Sick Baby, or Extramural Birth, Unknown/Other*)

**Admission type** (*Emergency, Urgent, Elective, Newborn, Delivery, Unknown/Other*)

**Payer class** (*Self-pay, Medicaid, Medicare, BC/BS, Commercial, HMO, Workman's Compensation, CHAMPUS/FEHP/Other Federal Government, Unknown/Other*)

---

<sup>5</sup> Comorbidity Adjusted Complication Risk – Brailer DJ, Kroch E, Pauly MV, Huang J. Comorbidity-Adjusted Complication Risk: A New Outcome Quality Measure, Medical Care 1996; 34:490-505.

**Discharge disposition** (*Home or Self Care, Short-term General Hospital, Skilled Nursing Facility, Intermediate Care Facility, Other Type of Institution, Home under Care of Organized Home Health Service, Left against Medical Advice, Discharged Home on IV Medications, Expired, Unknown/Other*)

Special risk factors are tailored to specific patient subpopulations. In the summary list above, these special risk factors include using birth weight instead of age for neonates, diagnosis detail, significant comorbidities, and defining procedures that depend on the disease stratum.

### **Functional Form and Risk Model Stratification**

Overall, the model follows a semi-log form (as specified in section I above), where the dependent outcome variable  $y$  is indexed for each outcome,  $l$ , at the patient level,  $i$ , for each provider,  $j$ , and principal diagnosis  $k$ . The vector of patient characteristics and socioeconomic factors is  $x$  and is similarly indexed such that the marginal effects of these independent variables on LOS depends on the principal diagnosis grouping. This diagnosis grouping method ( $k$ ) is based on the 3-digit ICD-9-CM principal diagnosis code. (The CareScience risk-adjustment methodology uses the 3-digit principal diagnosis code, modified to account for relative volume and clinical significance, to stratify the patient population into 136 categories that capture how patients present themselves to the hospital. Regressions are run for each of these strata.)

Separate regressions are run for each stratum to estimate the  $\beta_{kl}$  effect. The calibration program runs the regressions and generates the sufficient statistics, including the expected LOS (risk scores) for each patient and their standard errors.

### **Sampling for the Risk Model Calibration Dataset**

A representative sample of the patient population is used in the logistic regressions to generate the risk model's set of universal *beta* coefficients. The sample is drawn from all acute care facilities in the Premier client database, which spans 35 states and approximately 3.8 million discharge records. With a sample minimum of 5,000 cases per stratum, the database is statistically valid across outcome measuring, including in-hospital LOS.

### **Risk Model Calibration**

The nationally constructed sample is used to calibrate the patient LOS risk model. The calibration requires a separate regression to be run for each disease grouping. The resulting parameter estimates and corresponding estimated variance-covariance matrix are stored and provide the LOS risk model's universal *beta* coefficients. Application of these *beta* coefficients and their corresponding sufficient statistics to any set of patients will generate these patients' predicted LOS risks and standard errors.

The new set of risk scores for each disease grouping is compared against those generated from the individual client data. A series of econometric tests are performed to ascertain that the new set of parameter estimates (i.e., the universal *beta* coefficients) can be universally applied to current client data as well as other inpatient data. First, Spearman's rank correlation coefficient

between the individual client calibrated risk scores and the new set of risk scores is calculated for every outcome using the average deviation of a physician. Second, Pearson's product moment correlation coefficient between the two sets of risk scores is calculated for every outcome using the average deviation of a disease grouping. If these correlation coefficients have low values, a modified Chow test is performed to determine whether the sample is an accurate representation of the patient population. Otherwise, coefficients of variation are calculated to determine the stability of the parameter estimates across disease groupings.

Finally, the same econometric tests from above are performed on Medicare patients after they are risk-adjusted by the same set of parameter estimates.

### ***Data Sources – Public and private data***

The principal data sources are private Premier client data. To validate and supplement this basic data base, the Medicare Provider Analysis and Review (MedPAR) file from CMS is also used, as well as government-authorized data collected from every State that makes it available. The MedPAR file includes all Medicare inpatient data from U.S. states and territories for fiscal years with a two year lag. It is heavily populated with patients who are 65 and over, where younger patients would come under the disability provisions of Medicare or have end-stage renal disease. The files contain over 6,000 facilities and more than 12 million patient records. The State data includes all-payor inpatient data for from about 18 states (depending on the year), also with a two to three year lag.<sup>6</sup> It contains well over 2,500 facilities and over 20 million patient records.

Calibration of the risk model on the Premier client database instead of on public data presents the following advantages:

1. Public data is relatively outdated.
2. Public data release is subject to state-specific approval. States have been changing data use agreements on a regular basis in reaction to local and state political developments. The number of states releasing data to the public is unstable and increasingly unreliable.
3. States have been increasing restrictions on demographic data elements (such as zip code) and data on diseases with small volume in response to heightened patient privacy concerns and HIPPA requirements. This trend is expected to continue, again making complete access to this data set less likely over time.
4. Planned improvements to the risk adjustment methodologies require data that is not available in the public data. Premier has maximized the use of standard administrative data and sees little room for improvement without more detailed, electronically available clinical data.
5. Using the client data to calibrate the database allows us to update the calibration of the model on an annual/semi-annual basis as new clients/facilities begin using our methodologies.

Recall that Premier uses the calibration data set to determine the impact of patient characteristics on outcome variation (e.g., like the effect increasing age has on LOS for stroke patients). To

---

<sup>6</sup> Most recently the list of states include AZ, CA, FL, IA, MA, MD, NV, NJ, NY, OR, PA, RI, TX, VA, WA, and WI.

support this purpose, it is most important to have diverse data (a broad spectrum of patient types with sufficient observations in each) and accurate data in order to build a robust model of outcome risk assessment. Statistically and clinically, it does not matter how closely an individual hospital resembles the database globally, provided that the database encompasses hospitals that are similar. In this respect, as well, client data are superior to state data. In summary, Premier client data are more robust, reliable, complete, current, and significantly cleaner than the public data.

## **VII. Cohort Definition and Sampling Method**

### ***Cohort Definition***

As discussed the CareScience methodology can be adapted to any specific disease group, and is applicable across varying sizes of facilities and provider types.

### ***Sampling Method***

All patients meeting the specified diagnosis group definitions will be included in the analysis sets. No sampling of patients occurs. The entire patient population as defined by the specified diagnosis groups is used in the analyses.

## **VIII. ALOS Measure Reporting Format**

The CareScience ALOS measure is reported as the deviation of the observed in-hospital GM LOS from the expected in-hospital GM LOS.

### ***Interpreting Results: Statistical Significance***

To assist interpretation of ALOS comparison reports for a given case set, the statistical significance of each ALOS deviation (observed GM LOS – expected GM LOS) is provided. Deviations are given a “significance flag” that indicates whether the deviation could be plausibly interpreted as the result of pure chance. A double asterisk significance flag (\*\*) indicates 95% significance. Deviations with these double asterisk flags (\*\*) have a less than 5% probability that the deviation is due to chance. A single asterisk significance flag (\*) indicates 90% significance. Deviations with these significance flags (\*) have a less than a 10% probability that the deviation is due to chance. Large deviations tend to be significant, but not in all cases, especially when great uncertainty exists over expectations for the cases in the analysis set.

Statistical significance depends on the prediction error of the CareScience risk model. Prediction error, in turn, depends on how well the model fits the population on which it is calibrated. Hence, it is a characteristic of the model calibration, and not a feature of the group of patients in

the analysis set. Thus, as a practical matter, prediction error (hence, statistical significance) can be computed for any number of cases in the analysis set, even just one case.

The basis for computing statistical significance is to aggregate (as described below) the case-by-case prediction error. Just as the predicted outcome risk for each case is based on that patient's characteristics as processed by the CareScience model, so too does the model generate a prediction error for each case. The prediction error for the group of patients in the analysis set is derived by statistically aggregating the cases in the analysis set (with thousands, if not tens of thousands of patient observations within each stratum).

Aggregation does pose a challenge because it requires combining the uncertainty around the predicted value (risk) and the imprecision of the observed outcome value, especially when the number of cases in the analysis set is small. For this reason, one must be cautious in interpreting "significant" deviations when the number of cases in the analysis set is small. In such cases, the conclusion that the deviation is "significant" is based on an assumption that the observed rate does not derive from a highly idiosyncratic set of cases. This could be a problem when assessing individual physicians; but for larger samples, the law of large numbers helps make the assumption plausible.

### ***Aggregation***

The challenge arises when aggregating the standard errors to characterize any targeted grouping of patients. The expected value aggregated to the provider level for each outcome is the average of the patient level expected values for that provider or other grouping, DRG, MDC, etc. The estimated variance of the provider level estimator is derived from the regression statistics and the variance of the raw outcome measure is based on the sample.

As previously described, the deviation score represents the difference between the observed and expected values for ALOS. The provider-level deviation score is the average of the observed outcomes minus the average of the expected value. To determine the confidence interval around this deviation score, we must estimate the variance around the deviation score. The confidence interval allows us to determine if the deviation could be zero, indicating that there is not a significant difference between observed and expected outcomes.

All deviation scores will have an indicator of whether the deviation is statistically significantly different from zero. For example, a \*\* (95% significance level) indicates that there is less than a 10% probability that the deviation (expected - observed) is due entirely to chance. Hence, we can reject the null hypothesis of zero deviation with a 5% chance of a (type I) error.

## **IX. CareScience LOS Measure Purpose and Duration of Use**

The CareScience ALOS measure is designed for quality improvement and has been in use by Premier's client hospitals and health systems for 10 years. Our internal analyses have

demonstrated that increased risk adjusted hospital length of stay can indicate deviations from best practice due to special cause variation like complications etc. It is for this reason that high LOS can be a useful metric of quality. Care with high variation, care with high waste, and care with multiple special cause variation due to preventable causes is not high quality care, especially in the IOM framework of high quality care.

## **X. CareScience LOS Measure Audit Studies**

The CareScience risk-adjustment methodology has been audited through various statistical regression analysis tests. The methodology has also been audited by Premier's client hospitals and health systems through various studies.