Brief Overview of the Approach Used to Incorporate Advanced Statistical Considerations in the GIC CPII Round 5 Evaluation of the Quality of Care Delivered by Individual Physicians

---

VERSION 2.0 **2009.01.13**

Resolution Health, Inc. Proprietary Statement: This material constitutes proprietary and trade secret information of Resolution Health. This material may only be used for non-commercial purposes.

Resolution Health, Inc. | 10490 Little Patuxent Parkway, Suite 610 | Columbia, MD 21044 | 877-225-4744
www.resolutionhealth.com

Table of Contents

Purpose

The purpose of this document is to describe the statistical approach we are using in Round 5 of the CPII to provide a more sophisticated depiction of the quality of care delivered by a particular physician. The statistical approach we have employed not only takes into account several factors that could influence the estimate of the quality of care we observe for that physician, but also provides insight into the probability that the quality of care delivered by a physician is higher or lower than what we observe. There is more than one way one could take account statistically of the uncertainty around a point estimate of the quality of care delivered by a particular physician and more than one way one could use that information to place that physician into a particular performance "tier." In this paper we explain the statistical approach we have employed and the rationale for it and recommend an approach for placing a physician into a quality tier based on the output of our statistical model. We welcome feedback on our approach and anticipate improving it over time based on feedback and/or experience.

Introduction

During the first four years of the Group Insurance Commission (GIC) Clinical Performance Improvement Initiative (CPII), Resolution Health Inc. (RHI) produced a score of the quality of care provided by a physician based on the percentage of times the claims submitted by the physician showed that he/she complied with clinical practice guidelines that applied to his/her patients. For example, if MD_1 saw 20 patients during 2006, and two quality of care measures applied to 10 of these patients and three other quality of care measures applied to each of the other 10 patients, then there would have been 50 "observations" available regarding the quality of care provided by MD_1 in 2006. If the care provided by MD_1 was consistent with the relevant clinical practice guidelines in 40 of these 50 instances, then the quality of care score for MD_1 in 2006 would have been 40/50 = 0.80.

On several occasions over the past four years, the Massachusetts Medical Society and the health plans serving GIC beneficiaries pointed out that a score of this type for MD_1 constitutes a "point estimate" of the true quality of care provided by MD_1 and that there is some probability that the quality of care provided by MD_1 is higher or lower than what we observed. They also expressed concern that point estimates could be accurate, but nevertheless inadequate because, among other things, they failed to take into account patient non-compliance or the relative difficulty of complying with particular measures. The GIC and RHI agreed that there was some merit to those concerns and we have been working with expert biostatisticians from the Johns Hopkins School of Public Health to modify the quality of care scoring methodology we employ to reflect the statistical uncertainty regarding a point estimate of the quality of care delivered by any given physician.

Potential Sources of Uncertainty

Physicians deal daily with probability in their practices. For example, most tests that physicians order for their patients are not perfectly accurate. As a result, physicians recognize that there is some probability that test results for their patients may be higher or lower than their true values. Similarly, when clinical trials are reported in a journal article, the results of the study are typically reported as a "point estimate" with a 95% confidence interval around the point estimate which indicates the magnitude of uncertainty around the observed point estimate.

With respect to the CPII, there are many factors that may contribute to uncertainty in the quality of care scores that the GIC CPII produces:

1. We may only be observing the care a particular physician is providing to some of that physician's patients. For example, although the database we are using is extremely large, a physician may have patients who are members of a health plan whose data is not collected by the GIC. To the extent that the care provided by a physician is not always consistent with relevant clinical practice guidelines, the quality of care we observe in our database may not be an accurate reflection of the overall quality of care delivered by that physician.

2. The total number of observations that are available for a given physician influences the amount of uncertainty there is in an estimate of the quality of care delivered by that physician. The greater the number of observations that are available, the lower the uncertainty there will tend to be in the quality measurement that we produce.

3. Because of errors and omissions in coding on submitted claims, the data on which we base our measures are not always complete or accurate.

4. Because a widely agreed upon set of quality measures that addresses all of clinical medicine does not exist, our analysis is more limited in scope than we would like. The measures of quality of care that we have available are limited to those that are based on well-established clinical practice guidelines developed by clinical specialty societies and other highly regarded professional organizations, such as NCQA, AHRQ and NIH. Unfortunately, the quality of care measures that such organizations have produced do not reflect all the relevant types and aspects of care that a physician provides to his/her patients.

5. Certain clinical practice guidelines may be more difficult to comply with than others. An observed difference in the performance of MD_1 and MD_2 may reflect difficulty in complying with some measures as compared to others, rather than true differences in the quality of care delivered by those two physicians.

6. There may be differences in various characteristics of different physicians' patients, such as variation in patients' ability to pay for medications or tests, or

variation in patients' behavioral profiles, that influence the likelihood that those patients will comply with the care recommended by their physicians.

7. There may be a physician practice effect that results from aspects of the practice or setting in which a particular physician practices, such as IT infrastructure, nursing staff, the quality of a physician's partners, etc.

8. There may also be some effect from different care management or support programs sponsored by different health plans.

The Conceptual Approach We Have Employed

In Round 5 of the CPII, we have employed a statistical model that quantifies the uncertainty around a point estimate of the quality of care delivered by a particular physician (the "physician effect") while also attempting to control for the potential for any biases related to a) the mix of quality measures that applied to the physician's patients (the "measure effect"); b) the case-mix of a physicians' patients, each of whom has a particular likelihood of complying with his/her physician's recommendations (the "patient effect"); and c) the effect of the number of observations for a particular physician available in the GIC database (the "sample size effect").

The model's output is a probability distribution around a point estimate of the quality of care delivered by a particular physician, an example of which is shown in Figure 1 below. In this example, the point estimate of 0.7 for this particular physician's quality score is our best estimate of the chance that this physician will comply with relevant quality measures after controlling for the "measure effect," the "patient effect" and the "sample size effect." More specifically, 0.7 is the physician quality score with the highest probability in the posterior probability distribution that has been estimated based on the available observations of the physician's performance. However, in this example, as the probability distribution in Figure 1 shows, there is a 48% chance that the physician's "true" quality score is less than 0.68, the 20th percentile of the scores for all physicians, and a 19% chance that this physician's "true" quality score is greater than 0.74, the 80th percentile score for all physicians. The illustrative distribution below thus shows the best estimate of this physician's performance and the uncertainty in that estimate.

The more observations we have for a particular physician, the "tighter" the probability distribution will be around the best estimate of that physician's quality (i.e., the less uncertainty we have regarding that physician's quality of care). At first approximation, the width of the distribution, and hence the uncertainty in a physician's score, decreases proportionally to the square root of the number of observations available regarding that physician.[1]

The width of the probability distribution of the quality of care delivered by a physician depends not only on the *number of observations* available, but also on the *number of patients* that produce those observations for the physician. Suppose, for example, that MD_1 saw one patient to whom 10 quality of care measures applied and that the care provided by MD_1 was not compliant with any of these measures. Suppose further that MD_2 has the same 10 non-compliant measurements, but that MD_2 saw 10 different patients, and that one quality of care measure applied to each of those patients. Both MD_1 and MD_2 "scored" 0 out of 10. However, it is possible that MD_1 is an average physician and that the one patient for whom MD_1's observations are available has a higher than average tendency not to comply with a physician's recommendation. It is less likely that MD_2 is an average physician and that all of MD_2's 10 patients have a higher than average tendency not to comply with a physician's recommendations. As a result, even though there are 10 observations available for both MD_1 and MD_2, we are more certain about MD_2's score than we are about MD_1's score because there are more MD_2 patients considered in MD_2's assessment. The width of the probability distribution for MD_2's quality score will therefore be tighter than that for MD_1's quality score.

[1] For example, if one has 25 observations on a particular physician, one would need 75 additional observations (a total of 100) to halve the width of the posterior distribution (square root of 25 = 5 versus square root of 100 = 10). Similarly, if one has 9 observations on a particular physician, one would need 27 additional observations (a total of 36) to halve the width of the posterior distribution (square root of 9 = 3 versus square root of 36 = 6).

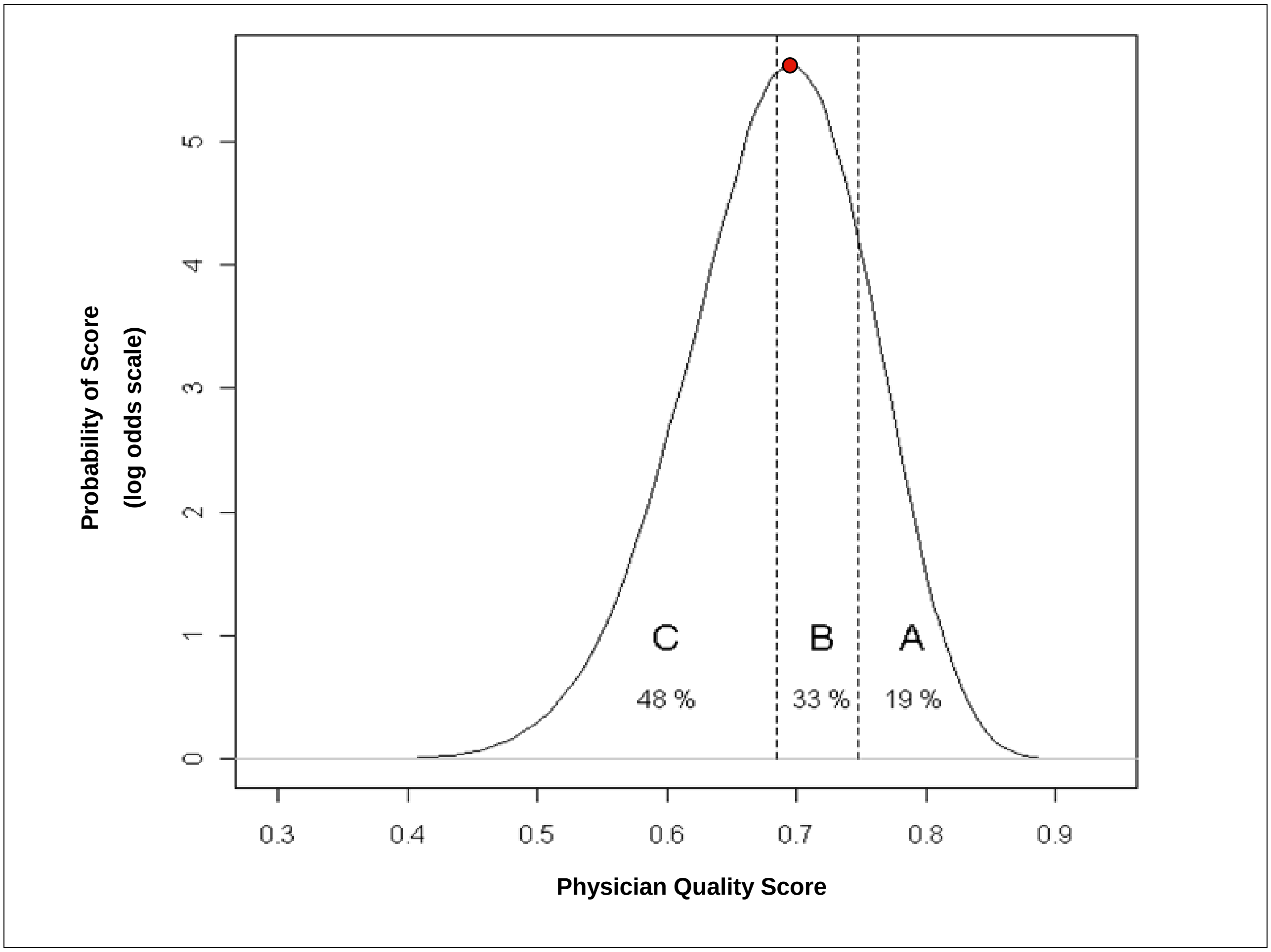

Figure 1: Sample Probability Distribution for a Physician's Quality Score

In summary, this sophisticated statistical model produces a distribution that depicts the probability that a physician's quality of care is a particular value or in a particular range, after accounting for the number of observations available for that physician, the types of measures that applied and the compliance of that physicians' patients compared to other patients. This approach to controlling for patients' ability and willingness to comply with a physician's recommendations begins to take "the patient effect" into account when assessing a physician's quality of care.

The Statistical Approach

The statistical model we have employed to produce these physician-specific probability distributions is a multi-level logistic regression model, also known as a "hierarchical" or "mixed" model (Diggle, Heagerty, Liang and Zeger, 2002). Multi-level models were developed more than 40 years ago (Rao, 1965). They have been used in many areas of science, as discussed by Diggle, et al (2002). They are now commonly used in hospital profiling (Christiansen and Morris, 1997; Shahian et al., 2005) and physician profiling (Hofer et al., 1999).

The model we have employed assumes that the probability (represented on a log-odds scale), that a physician complies with a particular set of quality of care measures that apply to a particular number of patients is a function of measure, patient, sample size and physician effects, the latter of which we use to quantify the physician's performance after adjusting for the other effects. [2] This model enables us to make adjustments for the number of observations available for a physician, and the variability among measures and patients, as opposed to calculating the simple compliance rate for each physician without adjusting for measure, patient and sample size effects.

Various Approaches to Place Physicians into Tiers

Over the past few years, the health plans that provide health coverage for the GIC have chosen to place physicians into tiers based on their assessment of the quality and cost of care delivered by those physicians. Although RHI has not developed or proposed a particular approach for assigning physicians to tiers, the results of our statistical analyses can be used to support a variety of approaches to do so.

There are two main approaches that could be used to define the tiers. In one approach, the tiers are defined to place a pre-specified fraction of physicians in each tier. For example, the "cut points" between tiers can be set to place the top 33.33% of scores, the bottom 33.33% of scores and the middle 33.34% of scores in tiers A, C and B, respectively. One can use this approach to define tiers based upon any percentile distribution one wants to employ. The alternative approach is to use absolute cutoffs (e.g., a quality score of 0.80 or above vs a quality score of 0.70 – 0.79 vs a quality score <0.70), regardless of the percentage of physicians that are categorized into each tier.

Regardless of which of these two approaches the GIC selects, and regardless of the tiering cut points that are established, health plans will need to employ a decision rule when deciding the tier into which a particular physician should be placed. For example, assume that the health plans establish three tiers -- Tiers A, B and C -- with cut points placed at the values indicated by where the dashed lines in Figure 1 above intersect with

---

[2] For computational feasibility, the contributing "effects" are idealized as following a normal distribution on the log-odds scale. This is a simplifying assumption that is commonly used to make computation of probability distributions feasible in hours rather than days or weeks. It is difficult to determine whether there is a significant departure from the normal distribution assumption given the sample size per physician. The impact of making the normality assumption diminishes, however, as the number of observations for a physician increases. Our convention of not producing a quality of care score for a physician with fewer than 10 observations thus reduces the impact of the normality assumption.

the X (horizontal) axis in the Figure. Once that decision has been made, several different decision rules could be used to assign a physician to a quality tier.

1. Assign the physician to the tier that corresponds to the peak (mode) of this physician's quality score probability distribution – in the case of the physician whose quality score probability distribution is shown above, the physician would be assigned to Tier B. As the figure shows, if this decision rule were employed, there would be a 48% chance that this physician's true quality corresponds to Tier C and a 19% chance that this physician's true quality corresponds to Tier A. In other words, the probability that this physician's true quality score is not in Tier B (48% + 19% = 67%) is higher than the probability that this physician's true quality score is in Tier B, even though the peak of his probability distribution falls within the Tier B range.

2. Assign a physician to the tier for which the *total* probability that the physician is in the tier is greatest. In the case of the physician whose quality score probability distribution is shown in the Figure above, this decision rule would result in the physician being assigned to Tier C, since the probability that the physician is in Tier C – 48% -- is higher than the probability that the physician is in Tier A (19%) or Tier B (33%). A drawback of this approach is that the physician is assigned to Tier C even though there is a higher chance (52%) that the physician is in Tier B or above.

3. Place the physician in Tier A if the probability that the physician is in Tier A is >50%, assign the physician to Tier C if the probability that the physician is in Tier C is >50%, and, in all other instances, assign the physician to Tier B. RHI recommends this latter approach since it eliminates the possibility that a physician would be assigned to either Tier A (or Tier C) at the same time that the probability that the physician is in a different tier or combination of tiers is greater than the probability that they belong in Tier A (or C). Another advantage of this decision rule is that it reduces the likelihood that a physician would be assigned to Tier C when they really are a Tier A physician and vice versa. A third advantage of this approach is that the physician's most likely score will be above the Tier A threshold (or below the Tier C threshold) if the physician is assigned to Tier A (C).

All of these approaches share the desirable property of providing the probability that a physician is in each potential tier, rather than providing only a single quality score value and an assignment to a particular tier.

2008 CPII Decision Rules to Place a Physician into a Quality Tier

In this year's CPII, the GIC has decided to categorize physicians into one of three quality designations (A, B, and C), with one third of physicians in each clinical specialty in the highest tier, one third in the lowest tier, and one third in the middle tier. Given this constraint of an equal percentage of physicians in each of the three quality tiers, physicians in 7 specialties – cardiology, endocrinology, rheumatology, OB-GYN, pulmonary medicine, pediatrics, and otolaryngology -- were assigned to a quality tier using the following decision rule:

1. In order to be assigned to the highest quality tier A, there had to be both 1) a 75% chance or greater that the physician truly belonged in tier A, based on the observations available for the physician and the results of our statistical model and 2) at least 30 quality of care observations available for that physician.

2. Similarly, in order to be assigned to the lowest quality tier C, there had to be both 1) a 75% chance or greater that the physician truly belonged in tier C, based on the observations available for the physician and the results of our statistical model and 2) at least 30 quality of care observations available for that physician.

3. In all other instances, the physician was assigned to the middle quality tier B for their particular specialty.

The minimum number of quality of care observations for physicians in 3 specialties – Hematology & Oncology, Neurology, and Ophthalmology – was lowered to 10. There still had to be a 75% or greater chance that the physician truly belonged to the highest tier or the lowest tier in order to assign that physician in the highest tier or the lowest tier respectively.

After a physician has been placed into a quality of care tier, the cost of care delivered by the physician relative to the cost of care delivered by all other physicians in the same clinical specialty is considered. Based on this cost of care (efficiency) analysis, each physician is placed into one of three efficiency tiers. Final tier assignments are based on both the quality of care and efficiency of care tiers.

Our expectation in final tier assignments is that, using this approach, approximately 20% of physicians in each specialty will be assigned to Tier 1 (the highest tier), about 65% will be assigned to Tier 2 (the middle tier) and about 15% will be assigned to Tier 3 (the lowest tier).

References

Christiansen CL, Morris CN. Improving the statistical approach to health care provider profiling. Ann Intern Med 1997;127:764–8

Diggle, PJ, Heagerty, PJ, Liang K-Y, Zeger SL. 2002. Analysis of Longitudinal Data. Oxford, Oxford University Press

Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physician “report cards” for assessing the costs and quality of care of a chronic disease. JAMA 1999;281:2098-2105

Rao, CR. The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. Biometrika 1965;52: 447-458

Shahian DM, Torchiana DF, Shemin RJ, Rawn JD, Normand S-L. Massachusetts cardiac surgery report card: implications of statistical methodology. Ann Thorac Surg 2005;80:2106–13