This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

TECHNICAL
REPORT

The Reliability of Provider Profiling

A Tutorial

John L. Adams

Prepared for the National Committee for Quality Assurance

RAND HEALTH

The research described in this report was prepared for the National Committee for Quality Assurance. The research was conducted in RAND Health, a division of the RAND Corporation.

Published 2009 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: http://www.rand.org
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Summary

Public and private purchasers and health plans are demanding more information about the quality and relative costliness of U.S. physicians to increase physician accountability and aid in value-based purchasing. Although performance measurement has been in place for some time in hospitals and managed care organizations (MCOs), the focus on physician profiling is a relatively new development. The inherent limitations of the available data at the physician level have brought to the fore technical issues that were less important at higher levels of aggregation in hospitals and MCOs. One of these technical issues is the reliability of a physician's performance measurement. Although a number of research efforts have evaluated quality measures for their reliability, the methods for doing so in practice may seem daunting to those designing and running performance measurement systems.

This report focuses on simple binary (pass/fail) measures. A patient may trigger one or more of these measures. These measures are then attributed to physicians, using various rules. HEDIS© measures are one example of the types of measures considered here.

Reliability is a key metric of the suitability of a measure for profiling because it describes how well one can confidently distinguish the performance of one physician from another. Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance.

There are three main drivers of reliability: sample size, differences between physicians, and measurement error. At the physician level, sample size can be increased by increasing the

number of patients in the physician's data as well as increasing the number of measures per patient.

For analysts not focused on the technical intricacies of physician profiling, the sudden emergence of reliability as an important property may come as a surprise. After years of focus on sample size, power, and confidence intervals, reliability has recently come into focus as a key issue. The reason this has occurred is the strong relationship between reliability and the key policy issues in the physician profiling debate. This is a logical consequence of the way that stakeholders are using physician profiling information.

There has recently been more interest in public reporting and pay for performance. The focus has been on putting physicians into categories. High-performance networks and three-star public reporting systems have emerged. A three-star system might give one star to physicians in the bottom quartile, two stars to physicians in the middle 50 percent of the data, and three stars to physicians in the top quartile.

Given the discrete categories that are being used, the big question becomes the probability of misclassification. If the categories are based on relative comparisons, reliability tells you most of what you need to know about misclassification in these systems. For a simple high-performance network system that flags a subset of the physicians as high-performing, there are two types of errors: (1) flagging a lower-performance physician as high-performance and (2) failing to flag a high-performance physician as high-performance.

In this report, we estimate reliability with a beta-binomial model. The beta-binomial is a natural model for estimating the reliability of simple pass/fail rate measures. There are also computational advantages to using the beta-binomial model, which is based on the beta distribution for the "true" physician scores. The beta distribution is a very flexible distribution on the interval from 0 to 1. The beta-binomial model assumes the physician's score is a binomial random variable conditional on the physician's true value that comes from the beta distribution.

This tutorial underscores that reliability is not just a property of a measure set but also depends what population is used to estimate the reliability. Whether a set of measures is useful for profiling providers depends on how different the providers are from one another. Measures that may be useful in one group of providers may not be useful in another group with little provider-to-provider variation. Similarly, as the providers under study increase their performance, the reliability may decrease if the provider-to-provider variance decreases over time. This is especially true as measures hit the upper limits of their ranges.

There are unanswered questions regarding how to calculate reliability for more complicated measures. Conceptual challenges remain, especially when multiple levels in the system may influence measure scores.