

THE EFFICIENCY SYMPOSIUM: METHODS, MEASURES AND POLICY IMPACT



EXECUTIVE SUMMARY

The emerging science of measuring physician efficiency utilizing episodes of care methodology took center stage during a half-day symposium conducted on September 16, 2008. Co-sponsored by the Wisconsin Collaborative for Healthcare Quality and the Wisconsin Medical Society, the symposium brought together providers

and purchasers of health care to gain a new understanding from two renowned experts in the field, Mark Rattray, MD and Elizabeth A. McGlynn, Ph.D. To access the PowerPoint presentations from the Efficiency Symposium, please visit www.wchq.org/events.

The Wisconsin Collaborative for Healthcare Quality (WCHQ) is a voluntary consortium of organizations learning and working together to improve the quality and cost-effectiveness of health care for the people of Wisconsin. The mission of WCHQ is to develop and publicly report measures of health care performance; design and promote quality improvement initiatives; and advocate for enlightened policies that support WCHQ's work.



Wisconsin Collaborative for
Healthcare Quality
PO Box 258100
Madison, WI 53725-8100
Web: www.wchq.org

- Mark C. Rattray, MD, is an expert on physician practice measurement. Dr. Rattray advocates for credible, sound health plan measurement processes and reliable depiction of practice performance information. Much of his time is spent educating healthcare stakeholders about practice measurement, and he has authored several articles on the topic. He founded CareVariance LLC in 2006 and serves as its President and Executive Consultant. Current clients include purchasing coalitions, health plans, physician groups and specialty societies, and regional health initiatives.
- Elizabeth A. McGlynn, Ph.D. is an Associate Director for RAND Health and holds the RAND Distinguished Chair in Health Care Quality. Dr. McGlynn is an internationally known expert on methods for assessing and reporting on quality of health care delivery at different levels within the health care system and has published extensively in the area. Dr. McGlynn is leading the COMPARE initiative which is developing a comprehensive method for evaluating options to improve the performance of the U. S. health care system. She is currently conducting research on the methodological and policy issues associated with implementing measures of efficiency and effectiveness of care at the individual physician level. She has been part of the Prometheus design team that has developed a new method of reimbursement for health care services.

THE EVOLUTION OF PHYSICIAN PRACTICE MEASUREMENT

Dr. Rattray examined the evolving measurement domains with a focus on process quality, resource use and costs, and episodes of care. The diverse measurement domains have resulted in program designs and policy interventions that initially held great promise for improving care, but revealed limitations and challenges when implemented.

Process Quality

In the 1990s employers, purchasers, and consumers began looking for a standardized way of measuring health plans and the care their contracted physicians provided. Process quality, best exemplified by the Healthcare Effectiveness Data Information Set (HEDIS) focused on measuring performance with respect to the delivery of certain preventative measures such as immunizations and screening for various types of cancer.

Resource Use and Costs

Resource use and cost measurement has generally taken the form of per capita expenditure measurement (i.e. how much is spent on a patient per year). Per capita expenditure measurement is best typified by the Dartmouth Atlas. Dartmouth Atlas research uses Medicare claims data to examine the care delivered to Medicare beneficiaries, providing comprehensive information and analysis on national, regional, and local markets. The research shows that most of the variance in the amount of care an individual receives is a function of the supply of services available in the community (e.g. number of physician specialists, Intensive Care Unit beds, and imaging centers) and not necessarily what is considered the most effective care.

In the highest spending regions of the country, more care is not necessarily better care. The Dartmouth Atlas studies show that in geographical areas with the highest costs there is a lower survival rate, a lower quality of life, and lower satisfaction with care.

Episodes of Care

A new measurement domain has recently emerged: episodes of care. Episodes of care is considered a “next generation” approach and has been defined as a series of clinically related health care events over a defined period of time, such as all claims related to a patient’s diabetes. The episodes of care methodology is viewed as a more logical way to view care delivery because it:

- Is patient- and condition-centric
- Recognizes that care is delivered over time
- Distinguishes between acute and chronic care

The episodes of care measurement methodology encompasses all the care delivered for a condition by examining health care claims, pharmacy claims, lab claims, and administrative data sets.

In a typical application, proprietary episode grouper software is used to combine a patient’s claims into an episode of care. Next, an actual dollar cost for that episode of care is computed by totaling the amounts of all of the claims for the episode. Using defined attribution rules, each episode is assigned to a particular physician. Physicians are then compared to the average costs attributed to physicians in the same specialty for the same type of care. Finally, a cost efficiency measure is calculated for each physician (e.g., ratio of actual costs to expected costs).



CONSTRUCTING PHYSICIAN-LEVEL EFFICIENCY METRICS

Dr. McGlynn introduced key considerations in whether the episodes of care methodology is ready for “prime time.” In other words, can the episodes of care methodology be used to accurately measure variations in the efficiency of individual physician performance using cost as the metric?

Dr. McGlynn began to answer the question by drawing on findings from a RAND research project in Massachusetts that was designed to (1) open up the “black box” and assess the methodology of profiling physicians, (2) identify and assess the many methodological choices that need to be made, and (3) highlight what the research findings mean for the development of public policy.

The RAND research team analyzed a data set of aggregated claims from four Massachusetts health plans. The data set represented three years of data (years 2003 – 2005), 2.9 million commercial enrollees (approximately 80% of the commercial market), and adults less than 65 years of age who were continuously enrolled for two years.

METHODOLOGICAL CONSIDERATIONS

Methodological choices must be carefully considered in order to build profiles that accurately measure physician performance. While there are a multitude of considerations, Dr. McGlynn focused on four primary issues that must be addressed before accurate measurement can result.

- Aggregation – Does combining data across health plans increase opportunities to accurately profile physicians?
- Attribution – Which physician is assigned responsibility for the care provided?
- Reliability – Is there a minimum reliability required for developing an individual physician’s profile?
- Classification – How are physicians assigned to categories of performance?

Dr. McGlynn identified the following issues and challenges related to the methodological considerations.

Aggregation

Aggregation of physician data across multiple health plans is critical to building accurate profiles, but is fraught with numerous challenges. These challenges include:

- *Identifying who is who.* Is physician A in one health plan, the same physician A in another plan? Does the master directory uniquely identify each physician accurately across multiple health plans?
- *Deciding on a physician’s specialty.* One health plan may identify Physician A as an internist while another identifies the same physician as a cardiologist. To what peer group are physicians assigned for

accurate comparisons among like-specialty physicians?

- *Putting together codes across health plans.* Health plans use diagnoses and procedures codes differently, and sometimes use home grown codes unique to their plan.

Aggregation is critically important to creating accurate physician profiles, but also very difficult. Dr. McGlynn warned that aggregation is detailed work, but essential to creating accurate physician profiles.

Attribution

Which physicians are responsible for the delivery and direction of care? Fundamental decisions must be made about how to attribute responsibility for care. These decisions include:

- Patient-based versus episode-based
- A plurality or majority of the costs versus a plurality or majority of the visits
- Assigning a single physician versus multiple physicians

Based on various combinations of these decisions, 12 different attribution rules can be employed. As Dr. McGlynn explained, there is no “right” approach, or gold standard when deciding what attribution rules to employ. In the end, the rules decided upon should be driven by the expected use of the information and who is using it. For example, if a consumer will use the information to choose a primary care physician, patient-based rules should be employed. If a consumer will use the information to pick a physician for a specific procedure, an episode-based rule would be best.

Reliability

How much of the variation in observed scores is explained by real differences in physician performance? When considering reliability here are some key points to remember:

- Reliability is an attribute of an individual physician’s own cost profile, not the measure overall
- The size of sample does not seem to

increase reliability for cost metrics, but does when using quality metrics

- Statisticians use a rule of thumb of requiring 90% reliability for drawing conclusions about individuals and 70-80% reliability for drawing conclusions about groups
- A significant fraction of individual physician profiles are unreliable – leaving a choice between not profiling most physicians or using sometimes less than reliable results to make decisions about physician performance

Classification

What method is used to assign physicians to different categories of performance? How is performance categorized as above average, average, or below average? Generally there are two methods used to categorize performance:

- Using cut points to assign a physician to a percentile band (<25th, 25th-75th, and >75th percentile)
- Statistical testing (i.e. do we have enough evidence that an individual physician’s performance is statistically significantly different than their peers?)

Physicians will be classified differently using these methods. A cut point system will identify some physicians as outliers when there is little statistical evidence that they are different from the average. Research on reliability demonstrates the importance of using more rigorous statistical tests when categorizing physician performance.

THE NEXT FRONTIER

While the episodes of care methodology holds promise, there are numerous challenges that must be addressed before it can be acclaimed as an accurate measurement of physician performance. Issues of aggregation, attribution, reliability and classification must be carefully addressed to assure physician performance is accurately measured.

“If it was easy, it would have been done already,” concluded Dr. McGlynn. Using episodes of care to measure individual physician efficiency is an emerging science and a work in progress. Dr. McGlynn does not imagine the process to be fast and emphasized the importance of having all stakeholders involved in the dialogue to move forward and develop better solutions.