

Episode-Based Physician Profiling: A Guide to the Perplexing

Lewis G. Sandy, MD¹, Mark C. Rattray, MD², and J. William Thomas, PhD³

¹UnitedHealth Group, Minnetonka, MN, USA; ²University of Washington, Seattle, WA, USA; ³University of Southern Maine, Portland, ME, USA.

Most current strategies to improve quality and efficiency in health-care delivery focus on measuring and improving physician practice. A new “second generation” of physician profiling—episode-based profiling—is moving beyond legacy “first-generation” physician profiles based on population health and preventive services measures. Episode-based profiling measures physician practice at the “episode of care” level with sophisticated analytic methods and tools using data from claim and other administrative data sets, and it has an underlying “theory of change” consistent with the evolution of the US health-care marketplace. While offering potential advantages in informing consumer choice and enabling practice improvement, episode-based profiling also has limitations and challenges, both analytically and in the process of physician engagement and improvement. Nonetheless, episode-based profiling is likely to continue to spread and have growing influence, and it has significant implications for research, policy, and clinical stakeholders.

KEY WORDS: profiling; quality improvement; transparency.

J Gen Intern Med

DOI: 10.1007/s11606-008-0684-z

© Society of General Internal Medicine 2008

INTRODUCTION

A series of analyses have documented significant opportunities to improve quality and efficiency in the US healthcare system,^{1,2,3} including recent reports from the Institute of Medicine⁴ that recommend a national, federal initiative to drive the health system's performance closer to the six aims outlined in “Crossing the Quality Chasm”: safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity.

Most of these improvement strategies focus on measuring and improving physician performance, as physicians' decisions have significant impact on the performance of the overall health system. This itself is not new; for several decades, organizations such as the National Committee for Quality Assurance (NCQA) and health plans have developed “first-generation” quality profiles, often based on population health and preventive service measures, such as HEDIS (Healthplan Effectiveness Data and Information Set). These first-generation profiles, while not without controversy, were conceptually and analytically straightforward, generally using a defined population as a unit of analysis. For example, to measure a

physician's rate of diabetic eye screening, all of the patients in a primary care physician's “panel” with diabetes served as the denominator, while the number with documented eye screening served as the numerator.

Recently, public and private purchasers have developed a new, “second generation” of initiatives that use “episode-based” profiling of physician practice to more directly target quality defects and the rising costs of healthcare services. This new approach typically performs data mining on claims, pharmacy, laboratory, and/or administrative data sets to construct patient “episodes,” where an episode of care refers to a period during which a disease process is present and is being managed, diagnosed, and treated by health-care providers.⁵ Using “episodes” as a unit of analysis requires sophisticated analytic tools and methods, and may seem non-intuitive and perplexing to many physicians. Yet because episode-based profiling provides the basis for new and emerging initiatives that can affect both physician practice and patient-physician relationships, it is critical that physicians understand how episode-based profiling works. Health services researchers and policymakers as well need to devote greater attention to the potential utility, limitations, and unintended consequences of these approaches.

CONSUMER-ORIENTED TRANSPARENCY AND A NEW “THEORY OF IMPROVEMENT”

Historically, common profiling efforts have been built using measures collected as part of health plan accreditation and performance reporting (e.g., HEDIS), clinical measures (process and/or outcome) for a defined set of procedures, e.g., NY State Coronary Artery Bypass profiling,⁶ or private, non-public physician feedback and data sharing focused on single specialties, such as the Northern New England Cardiovascular Cooperative⁷ and the National Surgical Quality Improvement Program.⁸

Despite the range and variety of these efforts, at present most physicians continue to lack any information on their own performance. A 2005 Commonwealth Fund report found only one-third of physicians had any comparative performance data available to them, with the most common source being health plans.⁹ Regardless of the profiling approach, most legacy profiling efforts had a “theory of change” that emphasized professionally oriented development and improvement.

Changes in the US healthcare marketplace, however, have catalyzed a “next-generation” approach to profiling with an alternative “theory of change” emphasizing physician “market share” shifts through informed consumer choice and/or direct financial incentives. As a result of the managed care backlash of the late 1990s, private health insurance increasingly migrated from closed-panel HMOs to “open access” PPO products. Consequently, consumers increasingly accessed

Received September 27, 2007

Revised February 21, 2008

Accepted May 15, 2008

care throughout a broad network, with no single group or primary care “gatekeeper” necessarily serving as a focal point of accountability. Recent growth of high-deductible “consumer-directed” health plans (CDHPs) has also increased the demand for publicly available information to make more informed choices regarding quality and costs. These marketplace changes have shifted the focus of performance measurement from the health plan to the individual physician.

OVERVIEW OF EPISODE-BASED PROFILING

Episode-based profiling is predominantly used in two ways to promote quality of care and/or cost efficiency improvement. First, by making performance information available to health plan members and giving the members financial incentives to utilize higher performing providers, health plans may seek to shift patient volume to physicians identified as providing better quality and/or more cost efficient care. Second, through pay-for-performance programs, health plans may attempt to promote performance improvement directly by giving financial incentives to physicians.

All episode-based profiling approaches rely heavily, if not exclusively, on administrative data resources—membership files, provider credentialing files, and claims records from commercial and government health insurers. Typically, all diagnostic and therapeutic services received by a health plan member—physician office visits, laboratory tests, imaging examinations, prescribed medications, office- and facility-based procedures, inpatient facility stays, physical therapy, etc.—result in claim forms being submitted to the health plan. Claims are submitted whether or not the plan, because of benefit limitations and design, ultimately pays all, some, or none of the service cost. A health plan claim usually includes three separate cost fields: (1) charge amount, the fee listed by the service provider as the charge for the service; (2) allowed amount, the amount contractually agreed upon by the health plan and the provider as payment for the service; and (3) paid amount, the amount actually paid by the health plan after subtracting patient payments for deductibles, co-payments, and co-insurance. In economic profiling of physicians, the allowed amount is generally used in cost efficiency calculations. Because claims data stored in electronic form are easily accessible for analysis, these databases represent unique resources for information on the physicians providing care to large numbers of patients. Some physicians flatly dismiss claims-based profiling analysis as incomplete and unreliable. While claims data sets have limitations, they are extensively used, and they have documented utility for measurement and improvement, and for answering research and policy questions.¹⁰

In a typical application, health plans process claims through “episode grouper” software, which aggregate each member’s claim records into defined “episodes of care” using data-mining algorithms. An actual cost figure is then calculated for each defined episode by summing allowed amounts of all claims included in the episode, including those for physician services, inpatient and outpatient facility services, prescription medications, and other services. After responsibility for each episode is attributed to a physician based on an attribution rule, such as “responsibility is assigned to a physician who accounts for 30% or more of professional and prescribing costs included in the episode,” an expected cost is calculated for each defined episode, usually as the average of actual cost of all episodes of the same type (e.g., all acute sinusitis episodes,

all type II diabetes episodes) managed by physicians in the same specialty. Next, for each physician, actual costs and expected costs of his or her attributed episodes are each summed, and the physician’s cost efficiency measure (e.g., ratio of total actual to total expected costs) is calculated. Physicians then are compared, within specialty, on relative cost efficiency performance.

Measures of quality performance are developed by using claims databases to extract information related to the degree of compliance with various clinical guidelines. Software implementing such performance criteria searches claims and membership files for “opportunities”—e.g., identifying patients with a diagnosis of type II diabetes—and then looks to determine whether the required procedure, medication, etc., was provided—e.g., the patient had at least two hemoglobin A1C tests in the last 12 reported months. Responsibility for quality opportunities [or “Quality Measurement Event” (QME), or a “case”] is then attributed to physicians using fixed attribution rules, such as “responsible physician is one in a relevant specialty (e.g., internal medicine, endocrinology) who had the largest number of Evaluation and Management (E/M) encounters during the profiling period,” and a compliance rate is calculated for each physician and each criterion by comparing opportunities with successes. Performance can be assessed in terms of relative compliance rate (e.g., 10% greater than average of the specialty) or as satisfying an absolute compliance rate standard.

STRENGTHS AND LIMITATIONS OF EPISODE-BASED PROFILING

Major advantages of episode-based profiling include administrative feasibility, minimal administrative burden for data collection, and the ability to compare performance against defined quality standards or specialty- or geographic-specific comparison cohorts. In addition, the “episode” view can be considered more “patient-centered,” as it analyzes care patterns and practices as experienced by patients across time and settings. On the other hand, episode profiling offers numerous opportunities for misidentification of high and low performing physicians. Variation in levels of detail and comprehensiveness of claims data across providers can introduce bias in comparative performance assessment. A physician’s cost efficiency performance is calculated using costs data from a sample of episodes managed during the profiling period. If episode responsibility is attributed incorrectly (i.e., to the wrong physician), if a few highly unusual and very expensive cases (i.e., cost outliers) distort estimates of physicians’ underlying cost performance, if some types of episodes considered in profiles are not representative of a physician’s usual practice, if risk adjustment is inadequate to control for effects of patients’ comorbid conditions on episode costs, or if the number of episodes available (sample size) for profile calculations is insufficient for reliable estimation, physicians’ cost efficiency scores may be inaccurate. Inclusion of hospital costs in episode-based profiles represents another challenge, as cost-efficiency measures can be heavily influenced by hospital costs, which are largely beyond physicians’ control, and may adversely impact profiles of physicians practicing in high-cost settings, such as academic medical centers.

On some key methodological issues, e.g., minimum number of episodes required for reliable profiles, definitive information

and standards are currently not available. In 2005, the Ambulatory Care Quality Alliance (AQA) and the NCQA launched an initiative to define methodological standards for episode-based profiling of physicians.¹¹

For quality performance measurement, the principal methodological issue is the limited number of quality criteria available. The RAND Corporation has identified more than 100 quality criteria that can be assessed using health-care claims out of approximately 1,000 guideline-based performance criteria.¹² Ingenix, which supplies a leading episode grouper used in efficiency profiling, offers a quality monitoring software package that includes 374 performance assessment criteria, and a number of other organizations provide software and services for physician quality performance assessment. Moreover, the underlying evidence base for guideline-derived performance criteria varies substantially. As of December 2007, over 100 ambulatory care measures have been endorsed by both the AQA and the National Quality Forum (NQF). But criteria having the NQF/AQA imprimatur relate to a relatively small set of clinical conditions—e.g., adult asthma, diabetes, coronary artery disease, congestive heart failure—and as a consequence are relevant to care delivered by only a few specialties.¹³ Attribution of responsibility for quality criterion opportunities and credit for criterion successes can also be problematic. As this field continues to evolve, consistent standards for use of episode grouper software applications, development of a wider range of truly evidence-based criteria, and uniform standards to support data aggregation initiatives to create larger sample sizes are important areas for ongoing development.

The recent IOM performance measurement reports note the current generation of performance measurement and transparency initiatives tend to focus predominantly on effectiveness, efficiency, and safety, with some attention to timeliness. Measuring and improving other dimensions of performance, notably equity and patient-centeredness, are not well developed to date, nor are measures of care coordination. Even within the scope of existing systems, limitations and unintended consequences should be kept in mind. For example, most health spending is concentrated among patients with complex chronic illness, and public reporting may have limited ability to influence consumer choice among such patients, as they already have established care providers. Moreover, public reporting could lead to “cream skimming”—improving measured performance by removing more difficult patients from one’s practice rather than through real improvement.¹⁴ Safety net providers may have patient populations and structural features that may disadvantage them in transparency programs. Finally, “report cards” may fail to distinguish between low- and high-“valued added” services, and may lead to further suboptimization of health care improvement efforts, unless weighting or other priority-setting approaches are used.¹⁵

PERFORMANCE IMPROVEMENT THROUGH PHYSICIAN ENGAGEMENT: A WORK IN PROGRESS

A number of national and regional health plans have developed and begun to deploy episode-based profiling programs, including Aetna, CIGNA, regional Blue Shield plans, and United-Healthcare.¹⁶ In theory, these initiatives can provide valuable information to both patients and physicians. As consumers have greater financial accountability (and risk) in their health

benefit plans, data to inform their care choices may result in better clinical and financial outcomes. Accurate profiling information can also be useful to physicians, as it can reveal how their own practice patterns may differ from colleagues.¹⁷

The process, however, of physician engagement and improvement is also a “work in-progress.” Competing priorities, variation in the clarity, sophistication, and sources of data for reports all can result in limited physician engagement. Physicians’ reactions to receiving comparative performance information can vary widely,¹⁸ and public reporting programs can generate controversy.^{19,20} Moreover, physicians may find profiling information acceptable for internal quality improvement efforts, while arguing the data are of insufficient reliability to allow public reporting of their performance.²¹

EPISODE-BASED PROFILING: IMPLICATIONS FOR PRACTICE, POLICY, AND RESEARCH

Limitations notwithstanding, public and private payers for health services are likely to continue to drive greater performance transparency in health-care delivery, using performance metrics that directly address well-documented variations in the quality and economic performance of the care delivery system. These tools and approaches, while imperfect, are likely to improve and become more standardized over time, and they have significant implications for practice, policy and research. First, for practicing physicians, these programs raise the importance of complete, accurate data collection and submission. Physicians who electronically submit accurate, fully documented claims to payers are more likely to have reliable episode profiles and will more likely receive “credit” for patient-level comorbidities and other risk-adjustment factors.

Second, physicians with better-organized and supported practice infrastructure are likely to be differentiated in their performance characteristics. In some instances, this will be directly attributed to an external organization’s “seal of approval,” such as NCQA Physician Practice Connection (PPC) recognition. In others, better scores may be enabled through the use of electronic health records, patient registries, internal reminders, and checklists, or other enhancements to practice organization.

Third, episode-based profiling is likely to catalyze the medical profession, particularly through specialty certifying boards, to accelerate the development of administrable, evidence-based performance measures. While initiatives such as AQA, NQF, and the AMA Physician Consortium for Performance Improvement (PCPI) are making significant progress, they will need to be augmented by accelerated efforts by specialty societies. One promising approach is to promote reciprocal relationships between episode-based profiling and specialty boards’ maintenance of certification programs, such as the program advanced by the American Board of Internal Medicine’s Practice Improvement Modules (PIM) program and many national and regional health plans.²²

Policymakers will need to carefully weigh the advantages and the limitations of episode-based profiling. In particular, private sector innovations can serve as a “laboratory” for learning how to best utilize episode-based profiling to advance policy goals.

The health services research community will need to devote greater attention to both the “basic science” of episode-based profiling as well as measuring the impact at the practice, patient, and community levels. At present, there is a signifi-

cant mismatch between the magnitude and speed of change in this arena, and the level of research effort.

Lastly, all stakeholders in the system need to engage in a collaborative process to improve the robustness, utility, and impact of episode-based profiling. Through greater understanding and focus by the practice, policy, and research communities, the pace of improvement in care, and the consequent value to patients and society, can accelerate.

ACKNOWLEDGMENTS: *Funded in part by the Robert Wood Johnson Foundation Health Care Financing and Organization Grant no. 60517 (Dr. Thomas). We are indebted to Steven A. Schroeder, MD, for his comments on a prior version of this paper.*

Conflicts of Interest: *Lewis G. Sandy MD: Employee/stock ownership/stock options: UnitedHealth Group; Mark C. Rattray MD: Employee: Booz Hamilton. Stock holdings: (<100 shares) T, ALU, AV, CMCSA, LST, NCR UNH (200 shares), WM (400 shares).*

Corresponding Author: *Lewis G. Sandy, MD; UnitedHealth Group, 9900 Bren Road East, Minnetonka, MN 55343, USA (e-mail: lewis_g_sandy@uhc.com).*

REFERENCES

- Institute of Medicine, Committee on Quality of Health Care in America. Crossing the quality chasm: A new health system for the 21st century. Washington, DC: National Academies Press; 2001.
- National Committee for Quality Assurance. The State of Health Care Quality 2006: Trends and Analysis. 2006 Sept [cited 2006 Dec 21]. Available from: <http://www.ncqa.org/tabid/447/Default.aspx> accessed May 15, 2008.
- McGynn EA, et al. The quality of health care delivered to adults in the United States. *N Engl J Med.* 2003;348(26):2635–45, Jun 26.
- Rewarding Provider Performance: Aligning incentives in Medicare (Pathways to Quality Healthcare Series). National Academy of Sciences. 2006.
- Hornbrook MC, Hurtado AV, Johnson RE. Health care episodes: definition, measurement, and use. *Med Care Rev.* 1985;42(2):163–218.
- Chassin MR. Achieving and sustaining improved quality: lessons from New York State and cardiac surgery. *Health Aff (Millwood).* 2002;21(4):40–51, Jul-Aug.
- O'Connor GT, et al. A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. The Northern New England Cardiovascular Disease Study Group. *JAMA.* 1996;275(11):841–6, Mar 20.
- Khoury SF. The NSQIP: a new frontier in surgery. *Surgery.* 2005;138(5):837–43, Nov.
- Audet AM, et al. Measure, learn, and improve: physicians' involvement in quality improvement. *Health Aff (Millwood).* 2005;24(3):843–53, May-Jun.
- Advancing Physician Performance Measurement: Using Administrative Data to Assess Physician Quality and Efficiency Pacific Business Group of Health available from: <http://www.pbgh.org/programs/PhysicianPerformance.asp> accessed May 15, 2008.
- AQA Alliance [homepage on the internet]. Washington, DC: AQA; c2005 [updated 2006 Nov 15; cited 2006 Dec 21]. Available from: <http://www.aqaalliance.org/> accessed May 15, 2008.
- RAND Health [homepage on the internet]. Santa Monica: RAND Health; c1994–2006 [updated 2006 April 03; cited 2006 Dec 21]. Available from: http://www.rand.org/health/surveys_tools/qatools/index.html accessed May 15, 2008.
- The National Quality Forum National Voluntary Consensus Standards for Ambulatory Care December 2007. available at: <http://www.qualityforum.org/projects/ongoing/ambulatory/index.asp> accessed May 15, 2008.
- Hofer TP et al. The unreliability of individual physician "Report Cards" for assessing the cost and quality of care of a chronic disease. *JAMA.* 281(22):2098–2105
- Hayward RA. Performance measurement in search of a path. *N Engl J Med.* 2007;356(9):951–3, Mar 1.
- Draper DA, Liebhaber A, Ginsburg PB. High-Performance Health Plan Networks: Early Experiences Issue Brief No. 111 May 2007 available from <http://www.hschange.org/CONTENT/929/#ib2> accessed May 1, 2008.
- Greene RA, Beckman H, Chamberlain J, Partridge G, Miller M, Burden D, et al. Increasing adherence to a community-based guideline for acute sinusitis through education, physician profiling, and financial incentives. *Am J Manag Care.* 2004;10(10):670–8, Oct.
- Beckman H, Suchman AL, Curtin K, Greene RA. Physician reactions to quantitative individual performance reports. *Am J Med Qual.* 2006;21(3):192–9, May-Jun.
- Ostrom CM. "Regence sued over ratings that cut 500 doctors" *Seattle Times* Sept 21, 2006. available at: http://seattletimes.nwsource.com/html/localnews/2003268477_regence21m.htmlv accessed May 15, 2008.
- Nakashima E. "Doctors Rated but Can't Get a Second Opinion" *Washington Post* July 25, 2007 A1 available from: <http://www.washingtonpost.com/wp-dyn/content/article/2007/07/24/AR2007072402545.html> accessed May 15, 2008.
- Lee TH, Meyer GS, Brennan TA. A middle ground on public accountability. *N Engl J Med.* 2004;350(23):2409–12, Jun 3.
- American Board of Internal Medicine [homepage on the internet]. Philadelphia: ABIM; c2004–2006 [updated 2006 Dec 12; cited 2006 Dec 21]. Available from: <http://www.abim.org/moc/healthcare/default.aspx> accessed May 15, 2008.