

**ECONOMIC PROFILING OF PHYSICIANS:
WHAT IS IT?
HOW IS IT DONE?
WHAT ARE THE ISSUES?**

A Guide Developed for the American Medical Association

June, 2006

**J. William Thomas, PhD
Institute for Health Policy
Muskie School of Public Service
University of Southern Maine**

INTRODUCTION

The purpose of this document is to enhance physicians' understanding of the process by which health insurance claim records are used to produce information intended to describe physicians' cost-efficiency performance. A large and increasing number of health plans now use provider-performance Web sites to make measures of physician cost efficiency available to health plan members, expecting that members will consult the data when selecting physicians for care. It is also increasingly common for health plans to use cost-efficiency data to partition physician networks into performance tiers, and then use differential co-payment or co-insurance rates to motivate health plan members to utilize physicians identified as cost efficient and to avoid physicians identified as cost inefficient.

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 co-payment, co-insurance, or deductible requirements, ultimately pays all, some, or none of the service cost. Health plans view their claim databases as comprehensive records of services received by patients. (Physicians, on the other hand, typically consider patient charts to be the only reliable source for such information.) Because claim data, stored in electronic form, are easily accessible for analysis, health plans regard the databases as ideal resources for information on the physicians providing care to their members.

The following steps are used by health plans to extract cost efficiency measures from their claim databases:

1. Claims are processed through “episode grouper” software, which aggregates each member’s claim records into “episodes of care,” 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.¹
2. An actual cost figure is calculated for each defined episode by summing allowed amounts² of all claims included in the episode, including those for physician services, inpatient and outpatient facility services, prescription medications, and other services.

¹ Examples of types of episodes include septicemia, acute bronchitis, viral meningitis, congestive heart failure, emphysema, and malignant neoplasm of the prostate. Episodes are defined for short duration conditions, such as acute bronchitis, and chronic conditions, such as congestive heart failure. At different times during the period covered by a claim database (e.g., one year) a patient might experience multiple episodes of the same type (e.g., viral skin infection episodes), and at any point in time the patient might be experiencing several different types of episodes (e.g., acute bronchitis, congestive heart failure).

² 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, allowed amount is used in cost efficiency calculations.

3. An episode expected cost is calculated for each defined episode, usually as the average actual cost of all episodes of the same type (e.g., all acute sinusitis episodes, all type II diabetes episodes).
4. Responsibility for each episode's actual and expected costs is attributed to a physician based on an attribution rule such as: "responsibility is assigned to the physician who accounts for 30% or more of professional and prescribing costs included in the episode."
5. Sums of actual costs and of expected costs are calculated for each physician based upon his or her attributed episodes.
6. A cost efficiency measure (e.g., ratio of total actual to total expected costs) is calculated for each physician, and physicians are compared, within specialty, on relative cost efficiency performance.

Although physician economic profiling analyses almost always include these general steps, final calculated results can differ significantly depending upon specific methodological details. In the following sections, we give an example of how cost efficiency might be calculated for individual physicians, and we define key terms and explain important concepts in economic profiling. Finally, we examine aspects of economic profiling methodology that can impact physicians' individual cost efficiency scores.

TERMS AND CONCEPTS

Efficiency and Cost Efficiency. Health economists define *efficiency* as a measure of resources required to achieve a given *level of outcome* – e.g., absence of pain, restoration of mobility. Consistent with this concept, the Institute of Medicine states that “in an efficient health care system, resources are used to get the best value for the money spent” (IOM 2001, page 52). However, when payers and purchasers speak of efficiency, they tend to focus on the cost of resources for a specified set of services, without explicit reference to outcomes. In September 2005, a conference convened by the Ambulatory Care Quality Alliance (AQA) and the National Committee for Quality Assurance (NCQA) proposed distinguishing between the terms *cost efficiency*, which would refer to relative resource utilization regardless of clinical outcome, and *efficiency*, which could refer to clinical resources used to achieve a measured clinical benefit (outcome).³ While we recognize that many still believe it inappropriate to discuss efficiency without explicit reference to outcomes, in this guide we will refer to *cost efficiency* of physicians; that is, we will be looking at relative costs or resource use among physicians when caring for patients without regard to the patients' outcomes.

Claim Database. Paid claim databases from commercial and government health insurers (The Center for Medicare & Medicaid Services) are the data source for assessments of physician cost efficiency performance. Claim databases maintained by health insurers, both commercial and government, include medical and, in many cases, pharmaceutical claim records for services provided to insured members during a specified period of time.

³ At a later meeting, AQA chose to apply the term “cost of care” to measures used in economic profiling.

Validity of Cost Efficiency Indicators. A valid indicator of cost efficiency is one that consistently and accurately distinguishes among efficient, average, and inefficient providers. Such measures can be characterized in terms of:

- *Bias.* A cost efficiency measure is systematically biased if it consistently shows a provider's performance to be poorer (or better) than is actually the case. What are the possible sources of bias in cost efficiency indicators? The most obvious is inadequate case-mix and risk adjustment. Case-mix adjustment controls for differences among physicians in the types of conditions treated. Risk adjustment controls for differences among patients in severity, comorbidities, and demographics. Demographic differences can be important because demographics often correlate highly with patient compliance with therapeutic regimens. If indicators are not properly adjusted to control for such differences when they exist, measured performance will be inaccurate.
- *Reliability.* Reliability refers to the likelihood that an indicator will yield the same value each time it is assessed. If an indicator is unreliable, it is likely to produce different views of cost efficiency performance, even when performance is unchanged and even though the indicator's value, averaged over a long period of time, might provide an accurate picture. The principal source of unreliability is the "small numbers problem." Indicators that are averages, such as average cost per episode, must be calculated from a large enough sample of cases to ensure that sampling variability (difference between sample mean and physician's true mean) is small. A second source of unreliability in cost efficiency indicators is the distorting effects of outlier cases. Various techniques, such as dropping outlier cases from analysis or truncating their values, exist for reducing the distorting effects of such cases.

EXAMPLE OF COST EFFICIENCY CALCULATIONS

Consider three cardiologists, Doctor A, Doctor B, and Doctor C, who are members of Alpha Health Plan's physician network. After Alpha's claims have been processed according to the steps described above, summary data shown in Table 1 have been compiled for the three physicians. For Doctor A, average actual episode costs are \$658 and average expected costs are \$548. Therefore, Doctor A's ratio of actual to expected cost is 1.20. Her costs probably would be considered by Alpha Health Plan to represent inefficient cost efficiency performance. Doctor B, with actual and expected costs equal to \$523 and \$564, respectively, has a ratio of 0.93, and Alpha would probably identify him as "average." Doctor C's ratio of actual to expected cost is $\$455/\$544 = 0.84$, identifying her as the most cost efficient of the three providers according to the ratio of actual to expected cost metric.

Table 1
Episode Costs for Three Cardiologists

Episode Type	Doctor A		Doctor B		Doctor C		
	Number of Episodes	Average Actual Cost	Number of Episodes	Episode Expected Cost	Number of Episodes	Episode Expected Cost	Average Actual Cost
Hyperlipidemia	1	\$613	12	\$520	1	\$520	\$385
Ischemic heart disease, except CHF, without AMI	6	\$1,293	32	\$1,308	3	\$1,308	\$1,685
Valvular disorder, without complication	2	\$1,151	9	\$622	0	\$622	\$0
Minor conduction disorder	8	\$712	21	\$577	5	\$577	\$321
Benign hypertension, without comorbidity	1	\$963	7	\$604	1	\$604	\$530
Cardiovascular disease signs & symptoms	16	\$440	53	\$357	9	\$357	\$265
Pulmonology disease signs & symptoms	3	\$521	18	\$315	0	\$315	\$0
Isolated signs, symptoms and non-specific diagnoses or conditions	3	\$128	21	\$143	2	\$143	\$120
Number of Episodes and Weighted Average Costs	40	\$658	173	\$564	21	\$544	\$455

KEY QUESTIONS ABOUT ECONOMIC PROFILING METHODOLOGIES

Are measures of cost efficiency performance related to measures of quality of care performance? If a physician performs every procedure, orders every test, and prescribes every medication that relevant clinical guidelines indicate should be performed, ordered, and prescribed, are episodes managed by that physician likely to appear more costly, and therefore *inefficient*, in economic profiles? Correspondingly, if a physician systematically fails to do all of the things that guidelines indicate should be done for patients in his or her care, are episodes managed by that physician likely to appear less costly, and thus relatively *efficient*, in economic profiles? A “yes” answer to either of these questions would indicate that measures of cost efficiency performance are confounded by measures of quality of care performance (i.e., good quality-of-care performance would be associated with poor cost efficiency performance), because degree of adherence to clinical guidelines is a definition of quality of care. As critical as these questions are, research has not yet been conducted to answer them. Several informal studies by health plans suggest essentially no relationship between cost efficiency and degree of compliance with process-of-care criteria. However, this conclusion is certainly not definitive. While a more formal study of this issue is currently underway, answers to the questions above are unlikely before 2007. Many in the provider community believe that until proof is available that they are not confounded by quality of care performance, cost efficiency measures should not be publicly reported or used by health plans in pay-for-performance or tiered network programs.

Should episode expected costs be risk-adjusted? Although economic profiling calculations control for case-mix differences among physicians by partitioning claim databases into episodes and then attributing the episodes to individual physicians, do these calculations also control for differences in severity, complexity, and demographic characteristics among patients managed? If calculations do not control for such factors, physicians could be motivated to avoid caring for sicker, more complicated, or more difficult to manage patients. To determine episode expected costs, most health plans simply calculate the average actual costs of all episodes of the same type. They assume that all episodes of, say, “ischemic heart disease, except congestive heart failure without acute myocardial infarction,” are similar, and that expected treatment costs of all patients having this condition are the same. But suppose a patient, in addition to this condition, is at the same time experiencing comorbid episodes of benign hypertension, type II diabetes, and diverticulitis? Won’t the cost of managing the patient’s ischemic heart disease be increased by the presence of these comorbidities? And thus won’t the physician who treats the ischemic heart disease be penalized in cost efficiency calculations by the assumption that all ischemic heart disease patients cost the same? Ingenix, vendor of the software package Episode Treatment Groups (ETGs), the leading episode grouper, also offers a person-level risk adjustment package called Episode Risk Groups (ERGs).⁴ The 2006 release of Ingenix’s ERG software, in addition to the currently available person-

⁴ *Person level risk adjusters*, which predict individual patient’s total health care costs during defined periods, are commonly used in calculating HMO capitation rates and in identifying patients for disease management programs.

level risk score, will provide episode level risk scores to account for effects of comorbidities on episode costs for individual types of episodes (e.g., ETGs). Thus, Ingenix, by decomposing its ERG person-level risk measure into episode level components, has concluded that episode costs can and should be adjusted for influences of comorbid conditions. The effectiveness of these risk adjusters and the degree, if any, to which they affect physicians' cost efficiency scores can be determined only after the new software release becomes available.

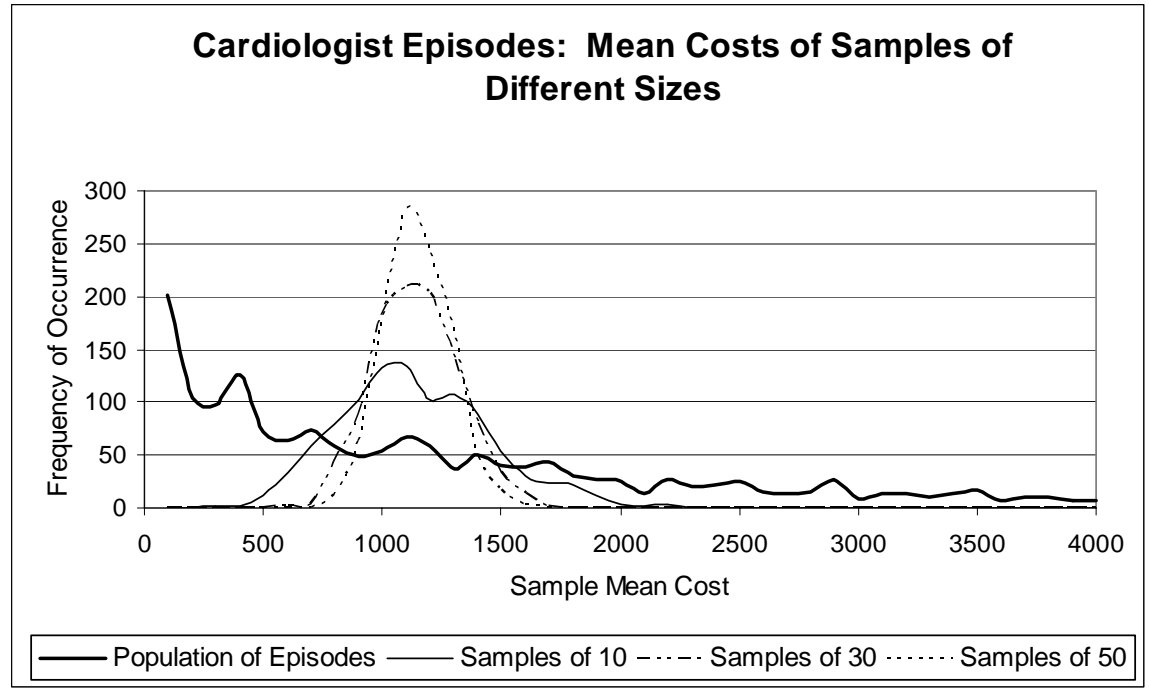
What are cost outlier episodes, why are they problematic in economic profiling, and what can be done to deal with them? Suppose a family practice physician treats 25 episodes of acute sinusitis, with costs ranging from \$75 to \$350 and averaging \$185. Suppose also that this physician manages a 26th episode, one that became complicated, and eventually accumulated expenditures of \$953. When the cost of this one episode is added, the physician's average cost for acute sinusitis becomes \$215, which is 16% higher than the average of the other 25 episodes. To minimize the potentially distorting effects of abnormally high or low cost episodes, it is common practice in economic profiling to identify cost outlier episodes and, within each episode type (ETG) in the database, either (a) trim (remove) outliers before making actual and expected cost calculations, or (b) truncate outlier costs to a specific percentage of the episode type cost distribution. For example, if actual costs for 95% of all acute sinusitis episodes in the database are below \$390, we could decide to truncate the costs of the remaining 5% to this level. If such an outlier procedure were followed in the example above, the family practitioner's average cost for acute sinusitis episodes would end up as \$193, instead of \$215. Research shows that choice of cost outlier methodology has an effect on reliability of physicians' cost efficiency scores, but that the effect is relatively small (Thomas and Ward forthcoming).

How is responsibility for individual episodes attributed to physicians? As noted in step 4 of the process described in the Introduction, a key task in economic profiling is attribution of responsibility for episode costs to individual physicians. If only one physician has participated in a patient's care, episode attribution is straightforward. However, if multiple physicians were involved, for example when a patient is seen initially by a primary care physician and subsequently referred to a specialist, how is responsibility for episode costs to be assigned among the involved physicians? A number of different rules, all *ad hoc*, are used by health plans. For example, one large national insurer attributes responsibility to each physician who had a professional fee included among the episode's claim. Another large national insurer attributes episode responsibility to the physician who had the most face-to-face contacts with the patient during the episode. A common practice is for health plans to attribute episode responsibility on the basis of the physicians' percentages of total professional fees, usually with some minimum percentage specified. Given the variety of episode attribution rules used, it is surprising that the conclusion from several research studies is that physicians' cost efficiency scores are not highly sensitive to episode attribution methodology (Thomas et al. 2004; Thomas and Ward forthcoming), because the vast majority of episodes are assigned to only one physician, regardless of the attribution rule used. For example, Thomas and Ward (forthcoming) found that when episodes are

attributed to physicians who account for 30% or more of an episode's professional and prescribing costs, more than 93% of cardiology episodes, 95% of family practice episodes, 99% of general surgery episodes, and 95% of neurology episodes are assigned to a single physician.

How many episodes are needed to construct a physician's economic profile? The factor having the single greatest effect on cost efficiency score reliability, and thus on the validity of physicians' cost efficiency scores, is the number of episodes used in profile construction. Figure 1 illustrates the concept of sampling variance and the potential effects of sampling variability on measured cost efficiency scores. The wavy line running from left to right near the bottom of the graph represents the cost distribution of approximately 1,600 episodes managed by a group of cardiologists in 2002. Episode costs range from under \$200 to greater than \$3,900; average cost of the population of episodes is \$1,093, and the standard deviation of the cost distribution is \$984. If a single health plan has only a few members served by this cardiology group, it will have only a few of the group's episodes in its claim database. The health plan's episodes can be viewed as a sample drawn randomly from the cardiologists' population of 1,600 episodes managed. From this sample, the health plan will calculate episode mean cost, use this mean as an estimate of the cardiologists' costs of treating health plan members and, after controlling for the types of episodes treated, this cost estimate will be compared to cost estimates for other cardiologists serving the health plan's members.

To generate the other lines shown in Figure 1, we repeatedly drew random samples of episodes from the cardiologists' population of 1,600, and then calculated the mean cost of the episodes in each sample. We selected 1,000 samples of 10 episodes each, and 1,000 samples of 30 episodes each, and 1,000 samples of 50 episodes each. For 10-episode samples, after calculating each sample's mean costs, we counted the number of samples having mean cost between \$0 and \$200, between \$200 and \$400, between \$400 and \$600, etc. We did the same thing for the 30-episode samples and the 50-episode samples, and then plotted the counts (frequencies) for each of the three different sample sizes. For samples of 10 episodes, a few of the sample means are less than \$400, and a few are greater than \$2,000. Ten percent of the 10-episode samples have mean costs that are less than \$700, and 10% have mean costs of \$1,500 or greater. Fewer than 1% of the 30-episode samples and none of 50 episode samples have mean costs lower than \$700; and only about 1% of 30-episode and 50-episode sample means exceed \$1,500. For 30-episode samples, 82% have mean costs that are in the range \$900 to \$1,300. This indicates that with a sample of 30 episodes, there is an 82% chance that episode mean cost will be within approximately \$200 of the cardiologists' true mean episode cost of \$1,093. With 50-episode samples, this probability improves to 92%, but with 10-episode samples the probability is only 58%.



Why do larger samples provide better cost estimates? Sample mean cost distributions are less variable with larger samples than with smaller samples because of a statistical relationship that says when samples are drawn from a population, the standard deviation of sample mean values is approximately equal to the standard deviation of the underlying population *divided by the square root of the sample size*. This relationship suggests that the 10-episode sampling distribution (the thin solid line in Figure 1) has a standard deviation of $\$984 \div \sqrt{10}$, or \$311. The actual standard deviation of the 1,000 10-episode samples is \$313. For the 30-episode and 50-episode samples, the relationship yields standard deviation estimates of \$180 and \$139, respectively, and the actual standard deviations for these sample distributions are \$178 and \$137, respectively. Because of this relationship, the accuracy of a health plan's estimate of a physician's cost efficiency score is greatly increased if a larger number of episodes is used for profiling. What is an appropriate minimum number of episodes? As important as this question is, little research has been done to answer it. Thomas (2005) suggests that the appropriate minimum differs among specialties. For cardiologists, 28% of physicians would be misclassified as inefficient with minimum episode samples of 10, while the misclassification rate with minima of 40 episodes would be 20%. For general surgeons, the misclassification rate would be 33% with 10 episode minima and 25% with 40 episode minima. For neurologists, misclassification rates would be 30% and 24%, respectively. These findings are considered tentative, because the analyses were done using data from a single, rather small health plan, and as yet no other researchers have published data for comparison.

There are three methods by which episode sample sizes can be increased to reduce misclassification. First, the time period over which episodes are accumulated can be extended. Some health plans, for example, base cost efficiency calculations on two years of claim data instead of one. The disadvantage of this approach is that cost efficiency performance information is less current, and therefore less meaningful. A second

approach is to analyze data for physician groups instead of individual physicians. Because the basic building block of economic profiles is the episode of care, profiles can easily be constructed for groups of physicians simply by aggregating the episodes attributed to individual physicians in the group. If group cross coverage practices routinely produce significant interactions between individual patients and multiple physicians within a group, a group profile may provide a more meaningful picture of practice patterns than separate profiles of individual group members. Finally, a third approach for increasing episode sample sizes is for multiple health plans in a geographic market to pool their health claim databases, thereby enabling each of the participating plans to profile a larger portion of its physician network than would otherwise be possible. In early 2006, The U.S. Agency for Healthcare Research and Quality, in conjunction with the Ambulatory Care Quality Alliance, initiated a set of pilot projects around the U.S. to demonstrate the utility of such pooling for enhancing the usefulness and accuracy of physician performance measurement.

IS ECONOMIC PROFILING HERE TO STAY?

What is the motivation for economic profiling? Because of constantly escalating health care costs, purchasers (employers) demand that their health plan contractors pursue any innovation, implement any program, adopt any change that might possibly help to control cost increases. The assumptions that underlie economic profiling include: (a) significant inefficiencies exist in the health care system, (b) providers who practice relatively efficiently can be distinguished from those who practice relatively inefficiently, and (c) appropriately targeted incentives can motivate providers toward more efficient use of health care resources and/or can motivate patients to select more efficient providers for their care. These assumptions are widely held, and physicians should expect to see more, not fewer, economic profiling programs in the future.

The goal of economic profiling, controlling costs by enhancing health system cost efficiency, cannot be achieved if profile information is inaccurate. Inaccurate measures of physicians' cost efficiency performance can damage reputations, and such measures will not produce efficiency improvements. While physicians' complaints about economic profiling will not make these programs go away, specific criticisms may encourage payers to take steps to improve information accuracy. And improved accuracy will benefit not only the physicians whose performance is being profiled, but also the payers themselves, since their goal is to promote cost efficiency improvement.

Currently, most health plans engaged in economic profiling have devised their own approaches for episode attribution, case mix adjustment, and the other issues described above. As a consequence, physician groups not infrequently find themselves rated highly on cost efficiency performance by some plans, and rated poorly by other plans for the same time period. In late 2005, the National Committee for Quality Assurance and the Ambulatory Care Quality Alliance initiated an effort to define methodology standards for economic profiling. When standards are eventually promulgated by these organizations (there is no current estimate for when this will occur), health plans are likely to adopt the recommended methodologies, and the value of cost efficiency information, both to health plans and to physicians, will improve. Nevertheless, until evidence becomes available to prove that better cost efficiency does not reflect

poorer quality of care; physicians may continue to distrust these measures as indicators of performance.

REFERENCES

Institute of Medicine (IOM). *Crossing the Quality Chasm*. (Washington, DC: IOM, 2001).

Thomas JW, Ward K, Grazier KL. Using profiling software to evaluate the practice efficiency of physician specialists. Final Report to the Robert Wood Johnson Foundation, Grant 047789. Portland, ME: Institute for Health Policy, Muskie School of Public Service, University of Southern Maine, 2004a.

Thomas JW, Ward K. Outlier treatment and episode attribution rules for economic profiling of physician specialists. *Inquiry* (forthcoming).

Thomas JW. Sample size considerations in economic profiling of specialist physicians. Portland, ME: Institute for Health Policy, Muskie School of Public Service, University of Southern Maine, 2005.