August 30, 2011

Ralph T. Hudgens, Commissioner
Office of Insurance and Safety Fire
708 West Tower, Floyd Building
Two Martin Luther King, Jr. Drive
Atlanta, GA 30334

RE: Physician Profiling Programs

Dear Commissioner Hudgens:

I am writing in my capacity as the president of the Medical Association of Georgia (MAG) to express my concern about the "profiling" and "pay for performance" metrics that health insurance companies in Georgia are using to rate physicians. As a bottom line, MAG does not believe that these are credible data points for cost or quality.

We asked Aetna and Humana to eliminate these measures from their physician profiles in letters that we sent to the companies in 2007 and 2010 (both attached), though neither complied. With that in mind, MAG’s House of Delegates passed a resolution at its 2010 meeting calling for you to prohibit insurance companies from profiling when it is based solely on "efficiency."

MAG genuinely appreciates the need to control medical costs, but we also believe that we need to employ systems that are fair and accurate. We are especially concerned that these profiling schemes continue to flourish despite the overwhelming evidence that suggests that the methods are flawed and unreliable.

The summary from a recent RAND Corporation study says that, "The majority of physicians in the data samples did not have cost profiles that met common thresholds of reliability and that the reliability of cost profiles varied greatly by specialty. In an illustrative two-tiered insurance product, a large fraction of physicians were misclassified as low-cost when they were actually not, or vice versa. Our findings raise concerns about the use of cost profiling tools, because consumers, physicians, and purchasers are at risk of being misled by the results."
The RAND findings have been highlighted in a number of publications, including the March 18, 2010 issue of the *New England Journal of Medicine* (Physician Cost Profiling – Reliability and Risk of Miscalculation*), the March 5, 2010 issue of *EMC – Health Services Research* (Incorporating Statistical Uncertainty in the Use of Physician Cost Profiles), and the May 19, 2010 issue of *Annual of Internal Medicine* (The Effect of Different Attribution Rules on Individual Physician Cost Profiles). We have attached copies of two of the articles for your information.

As the leading voice for physicians in Georgia, MAG believes that the profiling metrics that health insurance companies in the state are using to measure quality and cost of care are invalid and unreliable. We would, therefore, like to schedule a meeting with you to discuss ways that we can address this important issue.

As you know, it is imperative to maintain a favorable and sustainable practice environment if we hope to have enough physicians to care for patients in the state; addressing the profiling issue is an important step in the right direction.

MAG Executive Director Donald J. Palmisano Jr. will contact you in the near future to schedule a meeting. In the meantime, please feel free to contact Mr. Palmisano at 678-303-9250 or dpalmisano@mag.org in the event you have comments or questions.

My genuine thanks for your time and consideration.

Sincerely,

E. Daniel Deloach, M.D.
President

Enclosure

Copy: Donald Palmisano, JD, Executive Director, MAG
    John Goldman, M.D., Chairman, MAG Third Party Payer Committee
May 18, 2010

Thomas James III, M.D.
National Network Medical Director
National Network Operations
Humana, INC
National City Tower
12th Floor
Louisville, Kentucky 40202

Dear Dr. James,

I am writing on behalf of the Medical Association of Georgia (MAG) to ask you to eliminate your “Efficiency of Care” physician profile program in Georgia. The method that you are using to analyze physician costs is flawed and unreliable. What’s more, your low performance classification index will unfairly and inappropriately block physicians from the network.

MAG understands Humana’s desire to control costs and its focus on the physicians who write the orders for medical services. However, we wish to emphasize that physicians do not control the costs of hospitals and other ancillary services, nor do they have a lot of control over their referral choices. You assume physicians have the ability to make referrals to so-called “low cost” versus “high cost” facilities or services.

The majority of physicians are still in small, two-to-six physician group practices that serve a small geographic area surrounding their practice. With the consolidation of practice networks, hospital medical staffs, and contract agreements, physicians’ referral choices for hospitals, pharmacies and other ancillary services have become increasingly limited. Ironically, many of the so-called “high cost” hospitals used in physician’s cost profiles are in the very same networks with which health plans have contracted and require the patient and/or physician to use.

MAG encourages the use of physician data to benefit both patients and physicians and to improve the quality of patient care and the efficient use of resources in the delivery of health care services. While we respect innovations in assessing quality of care and cost efficiency, we do not believe the profiling methods that Humana uses are accurate and effective in achieving this goal. These concerns have recently been underscored by the American Medical Association (AMA) and by scientific studies published in The New England Journal of Medicine (i.e., the report on the Rand Corporation study, “Physician Cost Profiling – Reliability and Risk of Misclassification,” April 2010).
An excellent example of the flaws in your profiling model is demonstrated in the practice of Katarina Chiller, M.D., a dermatologist with Atlanta Skin Cancer Specialists (ASCS), who recently contacted MAG concerning its recent Humana profile. Outlined below is a list of the concerns that we have with your profiling approach, as described in your program materials provided to physician participants. The concerns are consistent with MAG and AMA principles and policies on tiered networks and with the New York Attorney General’s settlement agreement with CIGNA and other major health insurers, which provides profiling programs guidelines. The guidelines apply to any physician profiling program system that compares, rates, ranks, measures, tiers, or classifies a physician or physician group’s performance, quality, or cost of care against objective or subjective standards.

1. **Humana’s Efficiency of Care Profile** program title is misleading in that it does not measure efficiency and the program fails to have an independent oversight body. **Humana’s Profile is not voluntary and there is no indication of physician involvement in the conceptualization and development of the program.**

In reviewing Humana’s Efficiency Profile materials that were provided to the physicians at Atlanta Skin Cancer Specialists, we are first struck by the program’s misleading title. We believe that it should be more appropriately named a Relative Cost of Care Profile, since it is clearly focused on measuring only the costs of an episode of care, not the efficiency of care. Humana’s Efficiency Profile is not an “efficiency” profile in the common use of the word, in which the levels of resources or services are measured against the level of outcomes achieved. It is misleading both to physicians and patients.

In addition, we find no reference to the use of an independent oversight body to review your profiling process and results, a key requirement in the principles set forth by the AMA and the New York Attorney General’s settlement provisions. An independent and nationally-recognized oversight entity, not affiliated with the health insurer, is required to review the insurer’s programs that measure, report and tier physicians to ensure that the mechanism is fair and accurate. It is also necessary to ensure that all principles are being put into place, a step that is crucial in programs that greatly influence a physician’s future business and contracting abilities.

Another important principle which Humana fails to meet is that when data are collected and analyzed for the purpose of creating physician profiles, the methodologies used should be developed in conjunction with relevant physician organizations and practicing physicians. A physician’s determination of when, what and how medical services are provided are based on the medical needs of an individual patient and do not necessarily fit the same pattern as another patient. The physician’s expertise is vital in understanding these distinctions and other factors in developing a profiling method.

Profiling programs will have the greatest effect when they involve a broad base of physicians who are willing participants. Physicians want programs which are structured in a manner that is fair and understandable to them. They need to know and have confidence in the goals of a program and its method of operation. Physicians are not given the opportunity to voluntarily participate in Humana’s Efficiency of Care Profile since it is mandatory for all Humana contracted physicians who qualify.
2. Humana uses claims data which appears to be appropriately aggregated and reasonably timely, but fails to include any quality of care data or outcomes measures; Humana also fails to use any methods of medical record verification as a supplement for validating the claims data.

Claims data does not capture all of the information associated with a patient’s medical visit, therefore errors are inevitable. MAG believes that this is the missing capstone of economic profiling programs. Humana presently uses claims data that is appropriately aggregated across 26 specialties and timely to the profile period — semi-annually using a rolling 36-month period of claims data, currently between July 2006 and June 2009 for the March 2010 profile. The data still has almost a one-year lag period, with some data being outdated by two or three years, but it is consistent with other models currently in place. Ideally, MAG would like any physician profiling data to be based on concurrent data, which is timely and pertinent to the profile presented.

Importantly, the data gathered fails to include quality of care or other dimensions of care that would allow for a more fair and equitable assessment of the full value and worth of the medical care being provided. Foremost in profiling principles are that such programs should not profile physicians based on cost alone. As stressed earlier, in an open access product, such as a PPO, physicians have little or no control over the use of health care services and their associated costs by other health care providers.

In addition, Humana does not include any method of medical record verification as a supplement to further validate the claims data in its profiling. Multiple comorbidities and unresponsiveness to, or non-compliance with, treatment are just a few additional factors that would affirm why care may be more costly.

MAG cannot verify the accuracy of Humana’s data, although Atlanta Skin Cancer Specialists has not disputed the accuracy of the data contained in its profile report.

3. Humana’s measures used to rate physicians fail to contain:

   a. A sufficient data sample for arriving at an accurate profile
   b. An acceptable attribution method for determining what physician is most responsible for the costs of care
   c. Adequate risk adjustment components that take into account the entire patient factors that should be considered for valid results
   d. Profiling reports that are fully transparent, since all of the data used to arrive at the totals are not included in the reports provided to physicians
   e. Comparable physician peer groups for comparisons of episode of care costs in circumstances when there is a subspecialty

In brief, Humana uses Ingenix’s Episode Treatment Groups™ and Cave Consulting Group’s Marketbasket System™ consisting of algorithms and methodologies to create claims-based, longitudinal episodes of care at the patient level, which are formed using inpatient, outpatient, ambulatory and pharmacy claims data from the physician or physician group being reviewed. Cost efficiency is evaluated for all physician groups assigned at least 20 episodes. Humana assigns episodes of care costs to the physician groups that contributed to the patient’s treatment when the group was responsible for at least 25 percent of the professional costs. Our concerns with this approach are reflected below:
a. **Data Sample**: J. William Thomas, Ph.D., at the Institute for Health Policy at the University of Southern Maine, stated in “Economic Profiling of Physicians” that, “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.” He further states that in a study of the reliability of episodes for cardiologists with a sample of 30 episodes there was only an 82 percent chance that episode mean cost would be within approximately $200 of the cardiologists’ true mean episode costs. In comparison, Humana’s 20 episode benchmark falls even further below the 82 percent chance of being somewhat accurate with 30 episodes, and far below the 50 episodes of care which Dr. Thomas’ study showed was needed to achieve a 92 percent probability of accuracy. MAG would like to stress that the data sample should only cover the costs of service for which the physician is directly responsible.

Studies indicate that there is not a gold standard grouper system for looking at episodes of care. In the Rand Corporation’s April 2010 report on physician cost profiling they found that profiling results obtained from different applications using the data on the same population of patients may vary greatly. In fact, they found that the profile results were wrong up to two-thirds of the time for some groups and some 22% of physicians were misclassified in a two-tier system. (Adams, John L. et.al., *Physician Cost Profiling—Reliability and Risk of Misclassification*, Rand Health, April 2010)

In Humana’s profile of Atlanta Skin Cancer Specialists, many of the episodes listed for the 21 medical conditions were at zero or one. The two highest numbers of episodes at 13 and 25 were listed for benign neoplasm of skin and other malignancies of skin, respectively, only the 25 number being sufficient to base a cost profile.

Lumping hospital costs, pharmacy drugs, and other ancillary services into the physician’s claims data is extremely misleading and as stated earlier, are costs beyond the control of the physician. As you know, most physician practices are small businesses that would like to set their charges at amounts that allow them to cover their overhead costs and make a reasonable profit. Physicians, however, are often unable to establish their charges, but most often must accept the fee schedule offered by health plans. They have little negotiating leverage. Physicians order additional services based on the medical needs of their patient – services that are available within their geographic region and with the hospital or nursing homes to which they are most closely available – there is no incentive for them to do otherwise. To predicate physician cost profiles on ancillary services whose costs they have virtually no control and on payments established by health plan fee schedules for which they have little leverage appears foolhardy in the extreme. Such profiles misrepresent the true picture of costs for that physician’s services and are unfair and misleading to the patient.

b. **Attribution Method**: Humana assigns episodes of care costs to physicians/physician groups that are responsible for at least 25 percent of the professional costs. This may include physician-submitted claims for office visits, lab/path services, diagnostic tests, and medical and surgical procedures. Importantly, the 25 percent level established is purely random; it is faulty and has no scientific basis. It is decided on an ad hoc basis by each individual health plan. If only one physician has participated in a patient’s care, episode attribution is straightforward. However, if multiple physicians are involved in an episode of care there is
no easy or valid way of attributing the costs, bringing into question the entire model that’s in place.

c. **Risk Adjustments**: Humana states that it makes adjustments for a patient’s severity of illness, physician’s case mix, geographic location, as well as age, co-morbid conditions, and high and low-cost episodes, an improvement over earlier profiling methods. However, other risk factors also offer important elements that affect the cost of care, including health plan benefit type or socioeconomic status, and others that may affect access or adherence to care. In two independent research studies, numerous risk adjustment models fail to predict valid practice efficiency rankings and/or health care expenditures. (Thomas JW, et al., *Economic Profiling of Primary Care Physicians: Consistency Among Risk-Adjusted Measures*, Health Services Research 36:4, August 2004) (Martin K, et al., *Health-Based Risk Assessment: Risk Adjustment Payments and Beyond*, Academy Health, January 2004)

d. **Transparency**: Profiling principles state that profiling programs should be transparent, including disclosure of the methodologies, criteria, data and analysis used to evaluate physicians’ quality and cost-efficiency. Humana provides a five-page general description of its Efficiency of Care model and a one-page summary profile table to physicians being profiled with the total number of episodes, average cost, and average peer costs. It also separates the costs into professional, medical/surgical, ancillary, hospital inpatient, inpatient professional and pharmacy. It is not clear whether a more definitive picture of the actual data used is available but it is not presented as such. Humana is partly consistent with the principle of making their public reporting using rankings at the physician group practice level, in that they do use group ratings if that is the billing method used by the physician. If the physician bills with his individual ID number, Humana profiles him separately.

e. **Comparable Peer Groups**: Humana states that it measures each physician group’s cost efficiency to the comparable peer group for each applicable specialty. In Atlanta Skin Cancer Specialists’ (ASCS) case, Humana compared the group to other dermatology practices. The problem with this comparison is that ASCS exclusively treats cancerous skin conditions and the peer groups treat patients with a diversity of conditions not solely classified as cancerous, generally a more costly and intensely-treated service. In fact, in studying the Efficiency of Care Profile for ASCS, out of the 21 skin conditions listed for comparison, ASCS used almost exclusively two cancer diagnoses, benign neoplasm of skin and other malignancy of skin, which would skew any comparisons made with practices that were not skin cancer specialists.

4. **Humana’s profiling model fails to include a disclaimer that designations are intended only as a guide to choosing a physician, should not be the sole factor in selecting a physician, and have a risk of error; the method also fails to suggest that consumers should discuss designations with a physician before choosing him/her.**

5. **Humana’s profiling program information did not reference whether they offer a notice of use, change of designation, or program change, prior to disclosing the profile or introducing a new program. Humana does offer an appeal process, but does not indicate whether a quality and cost-efficiency rating is suspended when a physician makes a timely appeal or whether the physician has the right to submit additional information for an appeal.**
Clearly, the goal of economic profiling – controlling costs – cannot be achieved if profile information is inaccurate. Inaccurate measures of physicians’ cost efficiency performance can misrepresent a physician’s practice effectiveness and impugn his reputation with patients and, in the end, does not produce the efficiency improvements that the purchaser desires. MAG implores you to consider the evidence and make the necessary improvements in your profiling model to make it meaningful. We also urge you to avoid using the physician profiles to drop physicians from your network. Should you have any questions concerning this letter, please feel free to contact either of us or Camilla Grayson, MAG Director of Third Party Payer Advocacy, at 678.303.9275.

Sincerely,

John A. Goldman, M.D.
Chairman, Third Party Payer Committee

Gary C. Richter, M.D.
President

CC: Peggy Horne, VP Network Operations
Bob Barton, Hospital Contracting Executive
September 10, 2007

Ronald A. Williams, CEO
Aetna Inc.
151 Farmington Avenue
Hartford, Connecticut 06156

Dear Mr. Williams:

The Medical Association of Georgia (MAG) has received complaints from physicians across the state about the misleading and often damaging information included about them in the Aetna Aexcel designation program as published to patients. The purpose of this letter is to inform you about the nature of our concerns and to urge you to immediately remove the Aexcel designation program, as presently designed, as a show of good faith in your relationships with physicians.

The Aexcel program purports to allow patients to make well-informed health care decisions by informing them whether their physician practice has demonstrated effectiveness in the delivery of care, based upon whether they have met Aexcel clinical performance and efficiency measures and are therefore “Aexcel designated.” There is strong evidence that much of the profiling information used in Aexcel and other similarly designed programs are based on methodologies and measures, which are inadequate, inaccurate, and unproven. The goal of cost efficiency profiling, as we understand from your materials, is for controlling costs by enhancing health system cost efficiency. However, this cannot be achieved if profile information is inaccurate. Inaccurate measures of physicians’ cost efficiency performance can impugn physicians and damage reputations, and such measures will not produce efficiency improvements.

In fact, the American Medical Association passed a policy related to tiered networks such as yours. The policy states that third party payers should: 1) Disclose, in plain language, the criteria by which the carrier creates a tiered, narrow or restricted network, 2) Monitor their development so that they are not inappropriately driven by economic criteria and that patients are not caused health care access problems based on limiting specialists, 3) Seek legislation regulation which prohibits the formation of networks based solely on economic criteria, and 4) Insure that physicians are informed of the criteria for participating in those networks.

MAG’s concerns were first communicated to Dr. Jack Spicer during a meeting he had with MAG members and staff in April, when he outlined your plans to begin the program.
Our specific concerns are as follows:

1. Aexcel measures of cost efficiency performance do not appear to be directly tied to specific measures of quality of care performance, to which many plans are requiring conformance, making them arbitrary and isolated from the full range of care, which might be indicated for the condition. For example, if a physician performs every procedure, orders every test, and prescribes every medication that relevant clinical guidelines indicate should be performed, ordered and prescribed, episodes managed by that physician are likely to appear more costly, and therefore inefficient, in economic profiles. Unless this link is established, cost efficiency measures would not properly correlate with the clinical guidelines now being referenced by health plans.

2. Aexcel’s episodes of care do not appear to be sufficiently risk adjusted by severity, complexity, and demographic characteristics among the patients included, but are simply averaged by the actual costs of all episodes of the same type. These adjustments are necessary to reach an accurate and just conclusion. Ingenix, a prominent vendor for software package Episode Treatment Groups offers at least one person-level risk adjustment package called Episode Risk Groups, which indicates the added need for an episode cost adjustment for influences of co-morbid conditions.

3. It is not clear whether cost outlier episodes are identified and minimized to avoid the potentially distorting effects of abnormally high or low cost episodes. Research has shown that choice of cost outlier methodology has an effect on the reliability of physicians’ cost efficiency scores, however small. (Thomas JW, and Ward K. “Outlier Treatment and Episode Attribution Rules for Economic Profiling of Physician Specialists.” Inquiry.)

4. Claims data alone is not a sufficient indicator of the quality and cost efficiency of a physician’s patient care. As we understand, the Aexcel program uses the physician’s claims to aggregate each physician 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. Claims data represents only one element of the physician-patient treatment plan. It does not represent the entire patient and treatment profile, including the co-morbid conditions and other unique patient characteristics. Physicians consider the patient’s chart to be the only reliable source for such information. This alone calls into question Aexcel’s rating system.

5. Attribution of the cost of a patient’s episode of care to only one physician when multiple physicians have contributed to the costs is an inaccurate method of attributing costs in the rating system. As we understand, Aexcel’s responsibility for individual episodes are all attributed to only one physician, (the physician providing at least 30% of the professional and prescribing costs) even though there may be multiple physicians involved in the episode of care. On its face, this would not be an accurate cost accounting method.
6. Aexcel uses an insufficient volume of episodes in arriving at their cost efficiency score. Studies reveal that 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. Aexcel uses a volume of 20 episodes to construct a physician’s economic profile, an amount that falls on the low end for cost efficiency score reliability. Dr. J.W. Thomas of the Institute for Health Policy at the University of Southern Maine conducted a study on the sampling variability on measured cost efficiency scores for cardiologists based on the use of 10, 30 and 50 episode samples. Dr. Thomas found 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%. According to this study, Aexcel’s 20 episodes volume, alone, would create an accuracy rate somewhere between 58% and 82%, an unacceptably low rate of accuracy. (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)

The Medical Association of Georgia has a case in point concerning the questionable accuracy of Aexcel. Dr. John S. Harvey, a General Surgeon/ Trauma Surgeon in Alpharetta, Georgia was recently notified by Aetna that his practice did not receive Aexcel designation, nor will it be considered for Aexcel designation in 2008. Dr. Harvey contacted the Aetna office to inquire further about the basis for his exclusion and the specific methodology used. He was told that he would not be able to speak with the author of the letter concerning his non-designation, Mr. Ramzy ElGomayel, nor the Director of Georgia Network Market, which was also identified in Aetna’s letter to him. Only later, after suggested by MAG’s Third Party Payer Advocate to Dr Spicer, was Dr. Harvey finally contacted by Mr ElGomayel and Dr. Jack Spicer from your company.

After reviewing the back-up data used for his profile, Dr. Harvey noted that many of the costs included in the assessment were not his costs at all, but that of the hospital. Other service costs listed were often unclear, with additional costs often lumped in to the Physician Fee Schedule service payments at random and without clear explanation. A further failure of the program was to allow physician review of the findings before publication. We believe it is imperative that any program which presumes to report a physician’s clinical performance and efficiency should have review capabilities by that physician before the data is used and reported.

The Medical Association of Georgia has long supported improvements in clinical performance of its physician members. We were one of the first medical societies to organize an Institute for Excellence in Medicine several years ago, which presently oversees some 20 grant projects in this area. However, when health plans are using financial incentives in the management of health care, we believe they must be guided by carefully established principles, which do not compromise physicians or the medical care they provide, as outlined in our preface. These programs can have far-reaching effects and therefore must be scrupulously accurate in their execution.
Accordingly, on behalf of Dr. Harvey and other Georgia physicians who are the unwilling participants in Aetna’s Aexcel program, MAG asks that Aetna immediately suspend the Aexcel program and it’s physician designations as now published. Furthermore, we ask that you provide us a complete and indepth description of Aexcel’s program content and formula as used in determining a physician’s special designation and the rationale for including hospitals costs in an individual physician’s efficiency profile, when these are separate business entities of which the physician has no control. Moreover, the Medical Association of Georgia urges Aetna to promptly address the concerns expressed above by working to establish positive physician relationships based on accurate and valid clinical performance and efficiency measures. We also request that the Aetna Physician Advisory Board review our concerns.

Sincerely,

[Signature]

S. William Clark III, M.D.
President

[Signature]

John A. Goldman, M.D., Chairman,
MAG Third Party Payer Committee

cc: Troyen A. Brennan, M.D., M.P.H., Senior VP and Chief Medical Officer of Aetna
Robert Kropp, M.D., Senior Medical Director Southeast Region
Jack Spicer, M.D., Medical Director, Aetna of Georgia
John S. Harvey, M.D., Member MAG Third Party Payer Committee
Ronald M. Davis, M.D., President, American Medical Association
David A. Cook, Executive Director, Medical Association of Georgia
J. Leonard Lichtenfeld, M.D., F.A.C.P.
H. Robert Harrison, D.Phil., M.D., M.P.H.
Summary

This technical report accompanies the article "Physician Cost Profiling—Reliability and Risk of Misclassification," published in the New England Journal of Medicine (Adams et al., 2010). In this report, we provide more detail about the methods used to assess the reliability of physician cost profiling tools and the potential for misclassification of physician performance. We also present the results of our sensitivity analyses.

Purchasers are experimenting with a variety of approaches to control health care costs, including limiting network contracts to lower-cost physicians and offering patients differential copayments to encourage them to visit "high-performance" (i.e., higher-quality, lower-cost) physicians. These approaches require a method for analyzing physicians’ costs and a classification system for determining which physicians have lower relative costs. To date, many aspects of the scientific soundness of these methods have not been evaluated.

One important measure of scientific soundness is reliability. 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 that of 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.

The overall finding of the research is that the majority of physicians in our data sample did not have cost profiles that met common thresholds of reliability and that the reliability of cost profiles varied greatly by specialty. In an illustrative two-tiered insurance product, a large fraction of physicians were misclassified as low-cost when they were actually not, or vice versa. Our findings raise concerns about the use of cost profiling tools, because consumers, physicians, and purchasers are at risk of being misled by the results.
Incorporating statistical uncertainty in the use of physician cost profiles

John L Adams¹, Elizabeth A McGlynn¹, J William Thomas², Ateev Mehrotra³,⁴*

Abstract

Background: Physician cost profiles (also called efficiency or economic profiles) compare the costs of care provided by a physician to his or her peers. These profiles are increasingly being used as the basis for policy applications such as tiered physician networks. Tiers (low, average, high cost) are currently defined by health plans based on percentile cut-offs which do not account for statistical uncertainty. In this paper we compare the percentile cut-off method to another method, using statistical testing, for identifying high-cost or low-cost physicians.

Methods: We created a claims dataset of 2004-2005 data from four Massachusetts health plans. We employed commercial software to create episodes of care and assigned responsibility for each episode to the physician with the highest proportion of professional costs. A physicians' cost profile was the ratio of the sum of observed costs divided by the sum of expected costs across all assigned episodes. We discuss a new method of measuring standard errors of physician cost profiles which can be used in statistical testing. We then assigned each physician to one of three cost categories (low, average, or high cost) using two methods, percentile cut-offs and a t-test (p-value ≤ 0.05), and assessed the level of disagreement between the two methods.

Results: Across the 8689 physicians in our sample, 29.5% of physicians were assigned a different cost category when comparing the percentile cut-off method and the t-test. This level of disagreement varied across specialties (17.4% gastroenterology to 45.8% vascular surgery).

Conclusions: Health plans and other payers should incorporate statistical uncertainty when they use physician cost-profiles to categorize physicians into low or high-cost tiers.

Background

There is growing use of physician cost profiles which compare the costs incurred by a physician’s patients to an expected level of costs [1-10]. These profiles, alone or in conjunction with quality profiles, are being used by health plans for public reporting or for the creation of selective or tiered networks. In a selective network, patients can only visit low-cost physicians. In a tiered network, patients pay a smaller co-payment to see low-cost physicians. For example Aetna’s Aexecel network has two tiers of physicians and patients pay a $10-15 lower co-payment for visits with a low-cost, high-quality physician[5]. The goal of these initiatives is to create an incentive for physicians to decrease health care costs.

The physician cost profiles currently in use are based on commercial episode-grouper programs [2-9]. These programs group a patient’s claims into different episodes of care which comprise all services (e.g., visits, laboratory tests, hospitalizations) for a patient for a specific condition. This condition-specific approach in theory better addresses patient case-mix differences between physicians[11]. However, there is the concern that despite this advantage, cost data are still inherently "noisy" and a physician could be labeled as high-cost simply because of the combination of statistical uncertainty (“noise”) and insufficient sample size[1,5,12]. In statistical terminology this would be labeled Type 1 error.

To date health plans have generally not addressed statistical uncertainty when they identify physicians who are high, average, or low cost. Rather they use percentile cut-points across the distribution of physician cost profiles. For example, a health plan may label as low cost...
the 25\% of physicians at the bottom of the distribution [5]. Based on one author's (JWT) consulting experience, health plans do not use statistical testing because of the lack of an available method and the concern that statistical testing will identify so few low-cost or high-cost physicians to make selective or tiered networks impractical. Under statistical testing many truly low-cost physicians might be labeled as average (Type II error in statistical terminology) because the p-value threshold typically used (0.05) is stringent.

In this paper we introduce a method for conducting statistical tests in the setting of physician cost profiles. We then compare the two methods, percentile cut-offs and statistical tests, on how many high or low-cost physicians they identify and whether they agree on assignment. Finally, we discuss the implications of our results on current policy and the concern that statistical testing identifies too few low or high-cost outliers.

**Methods**

**Data sources and study population**

We constructed an aggregated commercial claims data set that included all professional, inpatient, facility, and pharmacy claims from four health plans in Massachusetts for 2004-2005. We analyzed all claims for the 1.13 million enrollees between the ages of 18 and 65 who were continuously enrolled for the two years. We used a unique physician identifier previously created by Massachusetts Health Quality Partners to link data from the four health plans at the physician level[13]. Our study population consisted of Massachusetts physicians who submitted at least one claim to one or more of the four participating health plans and were in a non-pediatric, non-geriatric specialty with direct patient contact. Pediatricians and geriatricians were excluded because our claims sample was restricted to patients 18-65 years old. More details on our study population are provided in Additional file 1.

**Constructing physician cost profiles**

Our methodology, which is described in greater detail in Additional File 1, was designed to replicate as closely as possible the cost profiling methods commonly used by health plans. It involved the following steps:

1. **Create standardized prices.** We averaged across the four health plans the mean allowed cost for each procedure, visit, service, or drug to create a standardized price for the state[14]. Before calculating the standardized prices we set all prices below the 25th percentile to the price at the 25th percentile and all prices above the 97.5th percentile to the price at the 97.5th percentile of the cost distribution, a process known as Winsorizing[15].

2. **Construct episodes of care.** We used Episode Treatment Groups\(^5\) (ETGs) software to aggregate each patient's claims into clinically related episodes of care (Version 6.0, Ingenix, Eden Prairie, Minnesota).

3. **Calculate each episode's observed costs.** The cost of each patient episode was calculated by summing the standardized costs of each service multiplied by the number of times the service was provided within the episode.

4. **Assign episodes to physicians.** The total cost of an episode of care was attributed to the physician who had billed the greatest fraction (minimum 30\%) of professional costs within the episode.

5. **Calculate "expected" costs.** For each type of episode (e.g., uncomplicated diabetes) the expected cost was the mean cost across all episodes attributed to physicians of the same specialty, for patients with the same level of co-morbidities. We used Symmetry's Episode Risk Groups\(^6\) to assign patient episodes to different co-morbidity levels.

6. **Construct composite cost profile.** We calculated a ratio based on all episodes attributed to each physician:

   \[
   \text{Composite Cost Profile} = \frac{\text{Sum of the Observed Costs}}{\text{Sum of the Expected Costs}}
   \]

**Creating standard errors for physician cost profiles**

Statistical uncertainty of a physician's cost profile is measured by the standard error, which will differ from physician to physician depending upon the specific set of episodes assigned to that physician. Calculation of the standard error must account for the number of episodes assigned, the characteristics of each episode type (e.g. minor skin inflammation vs. hyperlipidemia), and the effects of patient comorbidity levels (risk-adjustment). Details are provided in Additional File 1, but in brief we define the variance of a physician's cost profile as:

\[
\text{Var(Composite Cost Profile)} = \left( \sum \text{Var(Observed)} \right) \left( \sum \text{Expected Costs} \right)^2
\]

This assumes that the variance of the sum of the expected costs is small compared to the sum of the observed costs because it is based on all the data.

For the \text{Var(Observed)} we can use the variance of the entire population for each particular ETG and comorbidity combination. The standard error is the square root of this variance.

**Two methods for categorizing physician performance**

We compared two methods for putting physicians into the following three categories, low, average, and high cost. (We recognize that health plans sometimes only
use two categories, average vs. low cost [5], or average vs. high cost[16]. In the percentile cut-off method we rank ordered physicians based on their profile and labeled as low-cost and high-cost those in the bottom 25% and the top 25% respectively. We chose 25% as a cut-off as it is consistent with health plan initiatives [5,17]. In the second method we used a t-test to determine whether each physician’s cost profile is significantly different from the mean cost profile within their specialty. The t-test was created by taking each physician’s cost profile, subtracting the mean, and then dividing by the calculated standard error. In our primary analysis we used a p-value threshold of 0.05. In a secondary analysis we used a p-value of 0.20. We used a larger p-value to address the concern that statistical tests identify too few outliers (Type II error). As previously recommended by the National Committee for Quality Assurance we only profiled physicians with at least 30 assigned episodes[18].

We calculated how frequently the two methods disagree on the cost category assigned to each physician. We also measured disagreement using a weighted kappa. Under the weighted kappa method more weight is given to extreme disagreement (e.g., a physician labeled low cost under one method and high cost under the other) and less weight to less extreme disagreement (e.g., low cost under one method and average under the other).

All analyses were conducted in SAS version 9.1. (SAS Institute, Inc, Cary, NC).

Results
Our analyses are based on a sample of 8689 physicians in 27 specialties who had 30 or more assigned episodes. In our database, there was significant variation among specialties in median number of episodes assigned (range 50 for psychiatry to 680 for dermatology) [Table 1]. Across the specialties, the standard deviation of episode costs assigned was much larger than the mean episode costs (e.g., among episodes assigned to cardiologists, $2,300 mean cost, $13,322 standard deviation) which is indicative of high variance.

Based on the t-test with a p-value of 0.05, 25.2% of all physicians were placed in either the high- or low-cost category (range across specialties 5.1-66.2%) [Table 2]. Using a p-value of 0.20, 42.1% of all physicians were placed in either the high- or low-cost category (range across specialties 14.8% to 75.9%).

When comparing percentile cut-offs and t-tests (p-value ≤ 0.05), the two methods assigned 29.5% of physicians to different cost categories [Table 3]. The weighted kappa was 0.53 which would be interpreted as “moderate” agreement using the classification proposed by Landis[19]. Levels of disagreement varied across specialties, ranging from 17.4% for gastroenterology to 45.8% for vascular surgery. In supplementary analyses described in Additional File 1 we calculated the level of disagreement of t-tests and percentile cut-offs using an analysis where the fraction of high-cost and low-cost physicians is fixed across the two methods. In this supplementary analysis there was also substantial disagreement.

Discussion
There is growing use of physician cost profiles in an effort to decrease health care costs. Most of the policy applications of physician cost profiles such as selective or tiered networks require that physicians are divided into categories. Our results suggest that the method currently used by health plans to create these categories, percentile cut-offs, is inappropriate because it does not account for the substantial statistical uncertainty inherent in physician cost profiles. We demonstrate that when comparing the two methods, percentile cut-offs vs. statistical testing, almost a third of physicians are put into a different cost category. We advocate for the use of statistical testing in future cost profiling applications.

As noted above, health plans have been wary of using statistical testing because of concern that too few physicians will be placed into the low-cost or high-cost tiers and therefore making tiered networks impractical. In our study approximately 25% of physicians were put into the low-cost or high-cost category based on a t-test with a p-value cut-off of 0.05. Whether this is too many or too few depends on one’s perspective. Using a p-value of 0.05, minimizes Type 1 error (e.g., physicians who are not high-cost are misclassified as high-cost), but likely results in significant Type 2 error (e.g., physicians who are high-cost are misclassified as average). Physicians would prefer to minimize Type 1 error in this situation because being labeled as high-cost may damage their reputation as well as hurt them financially. Health plans are more likely to be concerned with Type 2 error. They hope cost profiles force physicians to focus on inappropriate spending. If a large number of high-cost physicians are misclassified as average, then the impact of the profiling efforts will be diminished.

There are alternative choices for statistical testing that might provide a better balance between these different viewpoints and their concern about the two types of statistical error. We illustrated one alternative using a p-value of 0.20 which likely increases Type 1 error and decreases Type 2 error. Under this alternative method, the number of outliers in our data increased from 25% to 42%. This might be a sufficient number of outliers from the perspective of a health
Table 1 Number of episodes and costs of episodes assigned to different specialties

<table>
<thead>
<tr>
<th>Specialty</th>
<th>Number of physicians</th>
<th>Median number of episodes assigned to physicians within specialty</th>
<th>Mean Costs per Episode ($)</th>
<th>Standard Deviation of Costs per Episode ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Specialties</td>
<td>8580</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allergy and Immunology</td>
<td>77</td>
<td>177</td>
<td>795</td>
<td>2936</td>
</tr>
<tr>
<td>Cardiology</td>
<td>474</td>
<td>92.5</td>
<td>2300</td>
<td>13322</td>
</tr>
<tr>
<td>Cardiothoracic Surgery</td>
<td>27</td>
<td>54</td>
<td>13488</td>
<td>34751</td>
</tr>
<tr>
<td>Dermatology</td>
<td>311</td>
<td>680</td>
<td>260</td>
<td>685</td>
</tr>
<tr>
<td>Emergency Medicine</td>
<td>521</td>
<td>97</td>
<td>708</td>
<td>2493</td>
</tr>
<tr>
<td>Endocrinology</td>
<td>97</td>
<td>116</td>
<td>925</td>
<td>2601</td>
</tr>
<tr>
<td>Family/General Practice</td>
<td>848</td>
<td>418.5</td>
<td>310</td>
<td>1447</td>
</tr>
<tr>
<td>Gastroenterology</td>
<td>356</td>
<td>292.5</td>
<td>1114</td>
<td>2683</td>
</tr>
<tr>
<td>General Surgery</td>
<td>407</td>
<td>142</td>
<td>2605</td>
<td>18993</td>
</tr>
<tr>
<td>Hematology/Oncology</td>
<td>122</td>
<td>93.5</td>
<td>2838</td>
<td>13420</td>
</tr>
<tr>
<td>Infectious Diseases</td>
<td>89</td>
<td>165</td>
<td>492</td>
<td>2160</td>
</tr>
<tr>
<td>Internal Medicine</td>
<td>2158</td>
<td>392.5</td>
<td>379</td>
<td>2317</td>
</tr>
<tr>
<td>Nephrology</td>
<td>111</td>
<td>67</td>
<td>1085</td>
<td>5131</td>
</tr>
<tr>
<td>Neurological Surgery</td>
<td>62</td>
<td>55</td>
<td>6936</td>
<td>18639</td>
</tr>
<tr>
<td>Neurology</td>
<td>249</td>
<td>108</td>
<td>1297</td>
<td>4605</td>
</tr>
<tr>
<td>Obstetrics and Gynecology</td>
<td>756</td>
<td>346</td>
<td>958</td>
<td>3554</td>
</tr>
<tr>
<td>Ophthalmology</td>
<td>421</td>
<td>326</td>
<td>358</td>
<td>913</td>
</tr>
<tr>
<td>Oral &amp; Maxillofacial Surgery</td>
<td>113</td>
<td>67</td>
<td>674</td>
<td>2097</td>
</tr>
<tr>
<td>Orthopedic Surgery</td>
<td>451</td>
<td>140</td>
<td>2008</td>
<td>9667</td>
</tr>
<tr>
<td>Otolaryngology</td>
<td>191</td>
<td>302</td>
<td>698</td>
<td>7164</td>
</tr>
<tr>
<td>Physical Medicine &amp; Rehab</td>
<td>70</td>
<td>24</td>
<td>1453</td>
<td>4586</td>
</tr>
<tr>
<td>Plastic Surgery</td>
<td>101</td>
<td>119</td>
<td>1150</td>
<td>3869</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>97</td>
<td>50</td>
<td>2320</td>
<td>3904</td>
</tr>
<tr>
<td>Pulmonary &amp; Critical Care</td>
<td>203</td>
<td>120</td>
<td>856</td>
<td>5555</td>
</tr>
<tr>
<td>Rheumatology</td>
<td>127</td>
<td>190</td>
<td>958</td>
<td>3958</td>
</tr>
<tr>
<td>Urology</td>
<td>191</td>
<td>242</td>
<td>1473</td>
<td>5547</td>
</tr>
<tr>
<td>Vascular Surgery</td>
<td>59</td>
<td>160</td>
<td>3223</td>
<td>14119</td>
</tr>
</tbody>
</table>

plan which is trying to create a tiered network. A second alternative, initially proposed by one of our co-authors, [20] is to rank order t-test values of physician efficiency measures and use a percentile cut-off (e.g. 20%) of this t-test distribution. This method has the advantage that a fixed percentage of outliers are identified for each specialty. The disadvantage is that for some specialties a high p-value threshold will be necessary to identify this fixed percentage of outliers. A high p-value threshold greatly increases the Type I error (e.g. physicians who are not high-cost are misclassified as high-cost). This method is described in more detail in Additional File 1. A third alternative is to reframe the statistical testing question. For example, the t-test could be reframed as the probability that the physician is different than the average physician in the lowest-quartile (as opposed to the average physician overall). These alternative approaches demonstrate that
Table 2 By specialty comparison of two methods for categorization and percentage of physicians identified as high cost or low cost

<table>
<thead>
<tr>
<th>Specialty</th>
<th>Number of physicians (≥ 30 episodes)</th>
<th>Physicians assigned a different cost category when comparing t-test (p = 0.05) and percentile cut-off (%)</th>
<th>Fraction of physicians identified as high cost using t-test N (%)</th>
<th>Fraction of physicians identified as low cost using t-test N (%)</th>
<th>T Test (p = 0.05)</th>
<th>T Test (p = 0.20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Specialties</td>
<td>8669</td>
<td>29.5%</td>
<td>1072 12.3%</td>
<td>1120 12.9%</td>
<td>1652 19.0%</td>
<td>2004 23.1%</td>
</tr>
<tr>
<td>Allergy and Immunology</td>
<td>77</td>
<td>27.3%</td>
<td>20 26.0%</td>
<td>19 24.7%</td>
<td>22 28.6%</td>
<td>30 39.0%</td>
</tr>
<tr>
<td>Cardiology</td>
<td>474</td>
<td>34.0%</td>
<td>51 10.8%</td>
<td>34 7.2%</td>
<td>65 13.7%</td>
<td>95 20.9%</td>
</tr>
<tr>
<td>Cardiothoracic Surgery</td>
<td>27</td>
<td>44.4%</td>
<td>2 7.4%</td>
<td>0 0.0%</td>
<td>3 11.1%</td>
<td>1 3.7%</td>
</tr>
<tr>
<td>Dermatology</td>
<td>311</td>
<td>21.2%</td>
<td>102 32.8%</td>
<td>104 33.4%</td>
<td>117 37.6%</td>
<td>119 38.3%</td>
</tr>
<tr>
<td>Emergency Medicine</td>
<td>521</td>
<td>34.9%</td>
<td>43 8.3%</td>
<td>37 7.1%</td>
<td>85 16.3%</td>
<td>70 13.4%</td>
</tr>
<tr>
<td>Endocrinology</td>
<td>97</td>
<td>33.0%</td>
<td>10 10.3%</td>
<td>10 10.3%</td>
<td>21 21.6%</td>
<td>15 15.5%</td>
</tr>
<tr>
<td>Family/General Practice</td>
<td>848</td>
<td>28.4%</td>
<td>102 12.0%</td>
<td>103 12.1%</td>
<td>173 20.4%</td>
<td>199 23.3%</td>
</tr>
<tr>
<td>Gastroenterology</td>
<td>356</td>
<td>17.4%</td>
<td>77 21.6%</td>
<td>59 16.0%</td>
<td>101 28.4%</td>
<td>88 24.7%</td>
</tr>
<tr>
<td>General Surgery</td>
<td>407</td>
<td>39.6%</td>
<td>23 5.7%</td>
<td>22 5.4%</td>
<td>39 9.6%</td>
<td>53 13.0%</td>
</tr>
<tr>
<td>Hematology/Oncology</td>
<td>122</td>
<td>36.1%</td>
<td>11 9.0%</td>
<td>7 5.7%</td>
<td>24 19.7%</td>
<td>26 21.3%</td>
</tr>
<tr>
<td>Infectious Diseases</td>
<td>89</td>
<td>32.6%</td>
<td>8 9.0%</td>
<td>13 14.6%</td>
<td>12 13.5%</td>
<td>18 20.2%</td>
</tr>
<tr>
<td>Internal Medicine</td>
<td>2158</td>
<td>24.6%</td>
<td>292 13.5%</td>
<td>382 17.7%</td>
<td>440 20.4%</td>
<td>625 29.0%</td>
</tr>
<tr>
<td>Nephrology</td>
<td>111</td>
<td>40.5%</td>
<td>6 5.4%</td>
<td>7 6.3%</td>
<td>14 12.6%</td>
<td>16 14.4%</td>
</tr>
<tr>
<td>Neurological Surgery</td>
<td>62</td>
<td>33.0%</td>
<td>7 11.3%</td>
<td>4 6.5%</td>
<td>11 17.7%</td>
<td>10 16.1%</td>
</tr>
<tr>
<td>Neurology</td>
<td>249</td>
<td>24.1%</td>
<td>35 14.1%</td>
<td>31 12.4%</td>
<td>56 22.5%</td>
<td>50 20.1%</td>
</tr>
<tr>
<td>Obstetrics and Gynecology</td>
<td>756</td>
<td>36.6%</td>
<td>54 7.1%</td>
<td>55 7.3%</td>
<td>91 12.0%</td>
<td>130 17.2%</td>
</tr>
<tr>
<td>Ophthalmology</td>
<td>421</td>
<td>18.3%</td>
<td>86 20.4%</td>
<td>89 21.1%</td>
<td>113 26.8%</td>
<td>128 30.4%</td>
</tr>
<tr>
<td>Oral &amp; Maxillofacial Surgery</td>
<td>113</td>
<td>32.7%</td>
<td>14 12.4%</td>
<td>7 6.2%</td>
<td>24 21.2%</td>
<td>18 15.9%</td>
</tr>
<tr>
<td>Orthopedic Surgery</td>
<td>451</td>
<td>39.7%</td>
<td>20 4.4%</td>
<td>29 6.4%</td>
<td>58 12.9%</td>
<td>76 16.9%</td>
</tr>
<tr>
<td>Otolaryngology</td>
<td>191</td>
<td>19.4%</td>
<td>30 15.7%</td>
<td>37 19.4%</td>
<td>46 24.1%</td>
<td>65 34.6%</td>
</tr>
<tr>
<td>Physical Medicine &amp; Rehabilitation</td>
<td>70</td>
<td>27.1%</td>
<td>8 11.4%</td>
<td>9 12.9%</td>
<td>12 17.1%</td>
<td>16 22.9%</td>
</tr>
<tr>
<td>Plastic Surgery</td>
<td>101</td>
<td>34.7%</td>
<td>10 9.9%</td>
<td>7 6.9%</td>
<td>20 19.8%</td>
<td>19 18.8%</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>97</td>
<td>30.9%</td>
<td>7 7.2%</td>
<td>13 13.4%</td>
<td>9 9.3%</td>
<td>25 25.8%</td>
</tr>
<tr>
<td>Pulmonary &amp; Critical Care</td>
<td>203</td>
<td>40.9%</td>
<td>15 7.4%</td>
<td>8 3.9%</td>
<td>27 13.3%</td>
<td>25 12.3%</td>
</tr>
<tr>
<td>Rheumatology</td>
<td>127</td>
<td>29.1%</td>
<td>17 13.4%</td>
<td>16 12.6%</td>
<td>24 18.9%</td>
<td>37 29.1%</td>
</tr>
<tr>
<td>Urology</td>
<td>191</td>
<td>31.9%</td>
<td>20 10.5%</td>
<td>17 8.9%</td>
<td>37 19.4%</td>
<td>43 22.5%</td>
</tr>
<tr>
<td>Vascular Surgery</td>
<td>59</td>
<td>45.8%</td>
<td>2 3.4%</td>
<td>1 1.7%</td>
<td>8 13.6%</td>
<td>2 3.4%</td>
</tr>
</tbody>
</table>

There are mechanisms to increase the number of outlier physicians while still addressing statistical uncertainty.

There are several important limitations to this research. We used a parametric approach to estimate the standard errors for the physician cost profiles. A non-parametric approach, for example using a null bootstrap estimator, might be more accurate. We did not use a bootstrap approach, because running bootstraps is often less intuitive for a policy and health plan audience and requires intensive computer resources. We felt a parametric approach would be more likely to be used in practice. Our analyses are based on data from four Massachusetts health plans. Although the specific level of disagreement will be different in other settings, we believe our overall finding of substantial
disagreement is likely generalizable. Lastly, this research does not address many other methodological issues that need to be addressed in developing cost profiles including the validity of the episode-grouper programs themselves and how care is assigned to a physician.

Conclusions
Using cost profiles to place physicians into high, average, or low-cost categories is becoming more common. We advocate that this categorization should address the statistical uncertainty inherent in profiles and we introduce a method of doing so.

Additional file 1: Technical appendix to accompany paper entitled "Incorporating statistical uncertainty in the use of physician cost profiles". The purpose of this technical appendix is to provide more detail about the methods in the manuscript in particular on how we calculated the standard errors of physician cost profiles as well as supplemental analyses on how we compare the two physician categorization systems.

Click here for file [http://www.biomedcentral.com/content/supplementary/1472-6963-10-57-51.doc]

Acknowledgements
We appreciate the input and support of Massachusetts Health Quality Partners who facilitated obtaining the health plan data and provided us with the physician directory. The research was supported by a contract from the U.S. Department of Labor. AM's salary was supported by a career development award (K23 RR024154-03) from the National Center for Research Resources, a component of the National Institutes of Health, and AM's participation in this research was supported by Grant #050517 from the Robert Wood Johnson Foundation's Health Care Financing and Organization (HCFO) program.

Author details
1 RAND, Santa Monica, CA, USA. 2Mayo School of Public Service, University of Southern Maine, Portland, ME, USA. 3Division of General Internal Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. 4RAND, 4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665, USA.

Authors' contributions
JA conceived of the idea and provided statistical oversight. JL, EAM, AMT, AE participated in the design of the study and writing of the manuscript. EAM conceived of the large project and obtained funding. AMT oversees the analyses and wrote the initial draft of the manuscript. All authors read and approved the final manuscript.

Competing interests
AMT has received consulting support on the topic of physician cost profiling from Agency for Healthcare Research and Quality, American Board of Medical Specialties, American Medical Association, Arkansas Medical Association, Blue Cross Blue Shield of Michigan, CHI-MIE Healthcare, Integrated Healthcare Association, Massachusetts Medical Society, Pacific Business Group on Health, and Wisconsin Medical Association. The authors have received a grant from the Massachusetts Medical Society, American Medical Association, and Physicians Advocacy Institute to study other aspects of physician cost profiling. None of the authors have any other financial interest in or a financial conflict with the subject matter or materials discussed in this manuscript.

Received: 26 May 2009 Accepted: 5 March 2010
Published: 5 March 2010

References


Pre-publication history
The pre-publication history for this paper can be accessed here: http://www.biomedcentral.com/1472-6963/10/57/prepub

Cite this article as: Adams et al. Incorporating statistical uncertainty in the use of physician cost profiles. BMC Health Services Research 2010 10:57.
The Impact of Different Attribution Rules on Individual Physician Cost Profiles

Ateev Mehrotra, MD, MPH, John L. Adams, PhD, J. William Thomas, PhD, and Elizabeth A. McGlynn, PhD
RAND Corporation (AM, JLA, EAM), Santa Monica, CA, and Pittsburgh, PA, Division of General Internal Medicine (AM), University of Pittsburgh, Pittsburgh, PA, and the University of Southern Maine (JWT), Portland, ME.

Abstract

Background—Health plans are profiling physicians on their relative costs and using these profiles to assign physicians to cost categories. Physician groups have questioned whether the costs category assigned to a physician is driven by the manner in which costs are attributed to physicians.

Objective—To evaluate the impact on physician cost profiles of 12 different attribution rules.

Setting—Massachusetts

Patients—1.1 million adults continuously enrolled in 4 commercial health plans in 2004 and 2005

Design—Using an aggregated database of claims from the 4 health plans, we created different cost profiles for each physician using 12 different attribution rules. The attribution rules differ on the unit of analysis (patient versus episode of care); signal for responsibility (costs versus visits); number of physicians that can be assigned responsibility; and threshold for assigning responsibility.

Measurements—Under each rule, we calculated the percentage of episodes assigned to any physician, calculated the percentage of costs billed by a physician included in his or her own profiles, and placed each physician into high cost, average cost, low cost, or low sample size categories. Compared to a commonly used default rule, we calculated what fraction of physicians are assigned to a different cost category using one of the other 11 attribution rules.

Address correspondence to: Ateev Mehrotra, RAND, 4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665, Phone: 412-683-2300 x 4894, Fax: 412-802-4972, mehrotra@rand.org.

Address for Reprints and Current Addresses:
- Ateev Mehrotra (reprints), RAND, 4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213. mehrotra@rand.org, Phone: 412-683-2300 x 4894; Fax: 412-802-4972
- John Adams, RAND, 1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407
- J. William Thomas, PO Box 9300, 509 Forrest Avenue, Portland, ME 04104-9300
- Elizabeth McGlynn, RAND, 1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138

Reproducible Research Statement
Protocol: Available to interested readers by contacting Dr. Mehrotra at mehrotra@rand.org
Statistical Code: Available to interested readers by contacting Dr. Mehrotra at mehrotra@rand.org
Data: Available through written agreements with the authors, Massachusetts Health Quality Partners, and the health plans who provided the data.

None of the authors have any other financial interest in or a financial conflict with the subject matter or materials discussed in this manuscript.
Results—Across the 12 different rules there was substantial variation in the percentage of episodes that could be assigned to a physician (range 20%–69%) and the mean percentage of costs billed by the physician that were included in the physician’s own cost profile (range 13%–60%). Compared to their cost category under the default rule, 17 to 61% of physicians would be assigned a different category across the 11 alternate attribution rules.

Limitations—Results might differ if data from another state or Medicare were used.

Conclusions—The choice of attribution rule affects how costs are assigned to a physician and can have a substantial impact on the cost category to which a physician is assigned.

INTRODUCTION

Health plan administrators and government payers are using the cost profiles of individual physicians for a variety of applications, including physician "report cards" and categorizing physicians into tiered products.(1) These applications are intended to generate incentives for physicians to decrease health care costs. Yet medical societies(2) and a state attorney general(3–5) have questioned the methodology used to create cost profiles, which compare a physician to his or her peers in terms of expenditures incurred.

One methodologic issue in creating the scores is how to determine which physician is responsible for a patient’s care when the patient sees multiple physicians.(6) Because there is no predetermined assignment of responsibility in most cases, analysts have developed algorithms to attribute responsibility on the basis of patterns of utilization found in data derived from health care claims. These algorithms are broadly referred to as attribution rules. One attribution rule, for example, assigns the care for a patient to the physician who accounted for the highest percentage of patient visits.(7,8) Another rule assigns the care to the physician who accounted for the highest percentage of the expenses incurred in caring for the patient.(2,9,10)

Various attribution rules have been proposed or used for physician cost profiling,(8,11) but little research has been conducted on the key question of whether the choice of attribution rule changes the cost profiles of individual providers. (12–15) To answer this question, we applied 12 different attribution rules (one default rule and 11 alternate rules) to an aggregated database of claims submitted to 4 commercial health plans in Massachusetts. Compared to the default rule, we assessed the impact of the alternate rules on (1) the fraction of care assigned to each physician, (2) whose care is assigned to the physician (care provided by the physician vs. care provided by his or her colleagues), and (3) the cost category (e.g. high cost, average cost) that the physician is assigned.

METHODS

Data Sources and Study Population

We obtained all claims (professional, facility, pharmaceutical, other) from four health plans in Massachusetts for 2004–2005. We used two years of data based on the recommendation of previous reports,(12,13) and we aggregated data across health plans to capture a larger share of each physician’s practice. The dataset includes claims from managed care, preferred provider organization, and indemnity products. We estimate that together the dataset included more than 80% of people with commercial health insurance in Massachusetts. Our analyses focused on the 1.1 million enrollees between the ages of 18 and 65 years who were continuously enrolled for the 2 years. More details on our criteria are available elsewhere. (16)
We included all Massachusetts physicians who submitted at least one claim to one or more of the health plans during the study period. To link data from the four health plans at the physician level, we used a physician identifier previously created by Massachusetts Health Quality Partners. We excluded pediatricians and geriatricians (to be consistent with our patient sample), physicians who did not have a specialty assigned, or were in a specialty without direct patient contact.

Physician Cost profiles and Physician Categories

Our methods for constructing physician cost profiles and physician categories were designed to closely follow or replicate the methods commonly used by health plans. Our methods are described in detail elsewhere. The steps are briefly outlined below.

Creation of standardized prices—For each service (visit, laboratory test, hospitalization, or prescribed drug), we examined the distribution of prices (reimbursement plus co-payment) across the 4 plans. We set all prices below the 2.5th percentile to the price at the 2.5th percentile of the distribution, and we set all prices above the 97.5th percentile to the price at the 97.5th percentile of the distribution, a process called Winsorizing. We then calculated the mean price for each service and assigned this standardized price to each service.

Construction of episodes of care—Each episode of care included the clinically related services (e.g., visits, laboratory tests, hospitalizations, prescriptions) delivered to a patient with a specific condition over a defined time period. To aggregate each patient's claims into episodes of care, we used Episode Treatment Groups (ETG)® (Ingenix, Version 6.0, Eden Prairie, Minnesota) which is a commercial product commonly used by health plans to group claims into episodes. We chose this commercial program over others because it is used by most Massachusetts health plans. It is also commonly used nationally.

The method by which the ETG grouper creates episodes is described in depth in previous publications. Briefly, the grouper takes all claims and places them into mutually exclusive and exhaustive categories. Each episode is marked with an ETG number. There are about 600 different types of episodes (ETGs) and examples include "hypo-functioning thyroid gland", "viral meningitis", and "cataract with surgery". Only certain types of claims can trigger an episode (e.g., evaluation and management visits, surgeries, hospitalizations).

We assigned episodes to a different comorbidity level using Episode Risk Groups (ERG) (Ingenix, Version 6.0, Eden Prairie, Minnesota). Under this system a patient's episode is assigned to a discrete risk level based on a retrospective risk-adjustment score based on patient demographics and co-morbidities. The number of risk levels varies by episode type and depends on the relationship between co-morbidities and costs observed among patients in the ERG development database.

Calculation of observed cost for each episode—We calculated the total cost of each episode (observed cost) by summing the standardized cost of each service multiplied by the number of times the service was provided within the episode.

Assignment of responsibility for care to physicians—We tested 12 different rules for assigning responsibility for patients or episodes of care, as described in detail below.

Calculation of the expected cost for each episode—For each episode, we calculated an expected cost which was the mean cost for all episodes attributed to physicians
of the same specialty (including those with low sample size) for patients with the same condition (ETG) and level of comorbidity.

**Construction of composite cost profiles**—We profiled physicians if they had 30 or more assigned episodes of care as previously recommended by the National Committee for Quality Assurance. For each physician their cost profile score was the total observed costs across all assigned episodes divided by the total expected costs. Therefore, the score is 1 if the observed costs equal the expected costs and it is >1 if the observed costs exceed the expected costs.

**Placement of physicians into cost categories**—We placed physicians into four categories: low cost, average cost, high cost, or low sample size (<30 episodes). Consistent with the method used by health plans, we examined the distribution of cost profile scores of physicians with 30 or more episodes and categorized physicians below the 25th percentile as low cost, those between the 25th and 75th percentile as average cost, and those above the 75th percentile as high cost. This was done separately for each specialty. In a sensitivity analysis we categorized the physicians using an alternative method, testing whether a physician’s cost profile was statistically different from the average physician within the same specialty.

**Description of Attribution Rules**

We created 12 different attribution algorithms that reflect the current range of rules being used or considered by payers (Table 1). Each rule is a combination of choices in the following four domains: unit of analysis (patient versus episode of care); signal for responsibility (professional costs versus number of evaluation and management visits); number of physicians that can be assigned responsibility (single physician versus multiple); and minimum threshold for assigning responsibility (majority of visits or costs versus plurality of visits or costs).

The first choice is whether to consider the costs incurred for all of a patient’s care or for each care episode. The patient-based rule would assign all of the patient’s costs to a single physician. The episode-based (condition-specific) rule assigns costs separately for each of the patient’s conditions to a different physician.

The second choice is whether to assign costs to the physician who accounts for the largest percentage of the total professional costs or to the physician who accounts for the largest percentage of evaluation and management visits. Based on convention we assigned costs based on only physician professional costs (e.g., reimbursement for visits or procedures) and therefore excluded testing or facility costs. The third choice is whether costs are assigned to a single physician or to multiple physicians. The fourth choice concerns the minimum percentage of costs or visits that needs to be reached before a physician is assigned the costs of care. If a 50% cutoff (majority) is used, the costs of care are assigned to a single physician. If a 30% cutoff is used, the costs are assigned to the physician with the largest percentage and at least 30% costs or visits (plurality) or to all physicians with at least 30% (multiple physician rules).

Because it is commonly used by health plans, we designated the episode-based costs plurality rule as the default rule.

**Analyses**

The analyses used descriptive statistics to compare the results of applying 12 different attribution rules (Table 1). For each patient, we used data from the two-year period to
calculate the costs and number of physicians involved. Similarly, for each episode, we used
data from that episode to calculate the costs and number of physicians involved. For each
rule, we calculated the percentage of episodes that could be assigned to physicians. We also
calculated the percentage of physicians who met the threshold for cost profiling (≥30
episodes).

To address whose costs are assigned to a physician, we examined the professional costs
billed by each physician and calculated the percentage of that cost incorporated into the
physician’s own cost profile under each of the 12 rules. We also looked at the converse—the
percentage of professional costs included in each physician’s profile that were actually
billed by that physician. Across all physicians we calculated the mean percentage and
standard deviation.

For each attribution rule, we used the results to assign each physician to 1 of 4 categories:
low-cost, average-cost, high-cost, or low sample size (<30 episodes). We compared the
default rule to each of the 11 alternate rules to determine the percentage of physicians for
whom the rules disagreed on the cost category assigned. For this analysis we excluded
physicians who were not assigned ≥30 episodes of care under any of the 12 attribution rules.

The funders had no role in the design and conduct of the study; the collection, management,
analysis, and interpretation of the data; or the preparation, review, or approval of the
manuscript.

RESULTS

Our study sample included 13,761 physicians who delivered 5,602,652 episodes of care for
1.1 million patients. The care of most patients and most episodes involved multiple
physicians. Among patients, 91% saw multiple physicians over the two-year study period,
and 61% saw five or more physicians. Among episodes, 54% involved multiple physicians,
and 9% involved five or more physicians.

Across the 12 different attribution rules, percentage of episodes that could be assigned to
physicians varied from 20–69%, and percentage of physicians with low sample size (≥30
episodes) from 39–66%. The mean percentage of a physician’s billed professional costs that
were included in the physician’s own cost profile ranged from 13–60% and the mean
percentage of professional costs included in a physician’s cost profile that were actually
billed by the physician range from 37% to 73%. (Table 2).

The analyses to determine disagreement in cost categories (Table 3) included the 9,741
physicians (71% of physicians in our sample) who had ≥30 episodes under any of the 12
rules. Rate of disagreement between the default rule and alternate rules ranged from 17% to
61%. In general, the highest disagreement rates were associated with patient-based rules
(47–61% range of disagreement) and multiple-physician rules (35–53%).

We tested the sensitivity of these conclusions to the manner in which disagreement was
measured and to the method of cost category assignment (Appendix). With one exception
there was only fair to moderate agreement across the different attribution rules when we
used a weighted kappa to measure agreement. Under this method we assigned different
weights to different levels of disagreement. When using statistical testing to assign
physicians to low and high-cost categories, disagreement rates were lower (11–50%) than
shown in Table 3 but still notable.
DISCUSSION

In creating profiles of the relative costs of care delivered by physicians, a number of methodological decisions must be made. We explored whether the choice of a rule for assigning costs to physicians affected the cost category to which physicians were assigned. We found that, compared to the most common rule used, 17% to 61% of physicians would be assigned a different category under an alternate attribution rule. Practically this means that if two health plans in a region chose different attribution rules, a physician will frequently be assigned a different cost category by the two health plans even if his or her care pattern was identical. Our findings are consistent with previous work that found that attribution rules applied to Medicare patients could affect the results of pay-for-performance programs. This result, however, diverges from what has been found in evaluating methods used to measure physician quality. In that literature, choice of attribution method has a relatively small effect on physicians’ quality profiles.

Our results help to explain why some physicians question cost profile attribution rules. No more than 60% of physicians’ billed costs were included in cost profiles under any rule. And of the physician costs assigned to a physician, up to two-thirds, were billed by other physicians. These findings might make physicians less responsive to efforts to use cost profiles to decrease spending. A related issue is what care assigned to a physician is truly “controllable” versus those costs that are driven by patient factors. To date, cut-offs such as 30% of the spending within an episode have been used as a method of determining assignment or control, but any such percentage cut-off is arbitrary and the appropriate cut-off may vary based on the condition being treated or the clinical scenario.

Given that the choice of attribution rule will lead to different conclusions being drawn about physicians’ cost performance, which rule is the best one? Unfortunately, there is no clear or simple answer to this question because “best” depends on what is important to each stakeholder and those perspectives vary. To illustrate this point we consider the views that might be held by a purchaser (health plan), a physician, and a patient. The purchaser is primarily interested in driving a change in the cost-related decisions of physicians, which will be easiest if the maximum number of physicians can be included in the profiling program. Using that criterion, the best rules include those that assign care to multiple physicians under which up to 66% of physicians having at least 30 episodes assigned.

From the physician perspective, the rule should accurately reflect what the physician is doing in practice. That means physicians might prefer profiles that capture a large proportion of their billed services and that reflect the care they billed rather than that billed by other providers. The best performing rule under the first criteria is the episode-based cost rule for multiple physicians which included on average 60% of a physician’s own costs in the profile. The best performing rule under the second criteria, that the profile does not include other physician’s care, is the episode-based, majority of costs, single physician rule in which 73% of the professional costs billed on average were by the physician being profiled.

From the patient perspective, the profiles should produce trustworthy information that is aligned with the decision the patient is being asked to make. Trust in the information will be undermined if the health plans in a region use different methods and different data because the different health plans will publicly report disparate results. Thus, patients are likely to be best served by efforts in which health plans pool their data and use consistent methods. Beyond that requirement, we might imagine two different types of choices being made by patients: choosing a primary care physician and choosing a specialist for consultation. In the case of choosing a primary care physician, the patient-based rules are most consistent with
the decision being made. When a patient chooses a primary care physician they are in some sense choosing that physician and his or her referral network. Under a patient-based rule the physician assigned a patient's costs is also assigned the costs provided by all the physicians caring for the patient. Alternatively, when choosing a specialist for a discrete reason, the episode-based rules are most consistent with the choice as the content of the profile is likely to be dominated by the types of services typically provided by those specialists.

There is no rule that best serves all perspectives. For this reason, transparency with respect to the methods used is critical. Purchasers will need to try and select a rule that balances these different perspectives. For example, in our study, under the episode-based cost multiple-physician rule, a high proportion of physicians were profiled and those profiles included a reasonably high proportion of physicians' own costs.

Our analyses have several key limitations. We used aggregated data from four commercial health plans. Patients from a single health plan typically comprise a small fraction of a physician's care. If we created cost profiles using a single health plan's data, fewer physicians could be profiled. It is also unclear how our results generalize to Medicare beneficiaries who generally receive care from a larger number of providers than commercially insured patients, a difference which could make attribution more difficult.(8) Although we tried to be comprehensive in our examination of attribution rules, there are variations in rules that we have not addressed. For example, some attribution rules use relative-value units rather than visits or costs as a signal for responsibility(22) and cutpoints other than 30% and 50%. (12,13,22) We also did not explore different attribution rules for different types of conditions.(31) It could be argued that the disagreement rates between attribution rules we report are overestimates. As we describe in our supplementary appendix, using statistical testing to categorize physicians results in lower, though still substantial, disagreement rates. Some health plans use two cost categories(22) instead of the three used in our analysis, and using two cost categories will obviously result in lower disagreement rates. Also, in some cases a large fraction of the disagreement between attribution rules occurs because of the difference in the number of physicians with a low sample size (<30 episodes). The cut-off of 30 episodes was based on previous recommendations.(21) If a higher or different threshold is used then our results would be different as fewer physicians would likely be included in any profiling effort.

The use of physician cost profiles has become more common. Our analyses emphasize that the choice of attribution rule affects how costs are assigned to physicians and that moving from one rule to another rule can make a difference in the cost category to which physicians are assigned. It is critical for health plans and others who create physician cost profiles to be transparent about how they assign costs to a physician. We hope these results prompt and inform a dialogue among stakeholders on which attribution rule should be used for different applications of cost profiles.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

**Primary Funding Source:** Department of Labor

Dr. Mehrota had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. We thank Barbra Rabson and Jan Singer of Massachusetts Health Quality Partners, who facilitated our access to the data sets used in this study.

Grant Support:

*Ann Intern Med.* Author manuscript; available in PMC 2010 November 24.
The research was supported by a contract from the U.S. Department of Labor (J-9-P-2-0033). Dr. Mehrotra’s salary was supported by a career development award (KL2 RR024154-03) from the National Center for Research Resources, a component of the National Institutes of Health, and Dr. Thomas’s participation in this research was also supported by Grant #60517 from the Robert Wood Johnson Foundation’s Health Care Financing and Organization program.

Dr. Thomas has received consulting support on the topic of physician cost profiling from Agency for Healthcare Research and Quality, American Board of Medical Specialties, American Medical Association, Blue Cross Blue Shield of Michigan, CIGNA Healthcare, Integrated Healthcare Association, Massachusetts Medical Society, Pacific Business Group on Health, Wisconsin Collaborative for Healthcare Quality, and the Wisconsin Medical Association. The authors have received a grant from the Massachusetts Medical Society and the American Medical Association to study other aspects of physician cost profiling.

References


Ann Intern Med. Author manuscript; available in PMC 2010 November 24.
32. Centers for Medicare & Medicaid Services. Berenson-Eggers Type of Service (BETOS) codes.
<table>
<thead>
<tr>
<th>Title of Attribution Rule</th>
<th>Unit of Care Attributed to Physician</th>
<th>Signal for Responsibility of Care&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Number of Physicians that Can be Assigned Care</th>
<th>Cutoff for Physician Assignment&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Episode, costs, plurality</td>
<td>Episode</td>
<td>Professional costs</td>
<td>Single</td>
<td>Physician responsible for plurality of costs</td>
</tr>
<tr>
<td>(Default Rule)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Episode, costs, majority</td>
<td>Episode</td>
<td>Professional costs</td>
<td>Single</td>
<td>Physician responsible for majority of costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Episode, visits, plurality</td>
<td>Episode</td>
<td>Evaluation and management visits</td>
<td>Single</td>
<td>Physician responsible for plurality of visits</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Episode, visits, majority</td>
<td>Episode</td>
<td>Evaluation and management visits</td>
<td>Single</td>
<td>Physician responsible for majority of visits</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient, costs, plurality</td>
<td>Patient</td>
<td>Professional costs</td>
<td>Single</td>
<td>Physician responsible for plurality of costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient, costs, majority</td>
<td>Patient</td>
<td>Professional costs</td>
<td>Single</td>
<td>Physician responsible for majority of costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient, visits, plurality</td>
<td>Patient</td>
<td>Evaluation and management visits</td>
<td>Single</td>
<td>Physician responsible for plurality of visits</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient, visits, majority</td>
<td>Patient</td>
<td>Evaluation and management visits</td>
<td>Single</td>
<td>Physician responsible for majority of visits</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Episode, costs, multiple physicians</td>
<td>Episode</td>
<td>Professional costs</td>
<td>Multiple</td>
<td>All physicians responsible for ≥30% of costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Episode, visits, multiple physicians</td>
<td>Episode</td>
<td>Evaluation and management visits</td>
<td>Multiple</td>
<td>All physicians responsible for ≥30% of visits</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient, costs, multiple physicians</td>
<td>Patient</td>
<td>Professional costs</td>
<td>Multiple</td>
<td>All physicians responsible for ≥30% of costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient, visits, multiple physicians</td>
<td>Patient</td>
<td>Evaluation and management visits</td>
<td>Multiple</td>
<td>All physicians responsible for ≥30% of visits</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>The definitions of professional costs and evaluation and management visits are available in the appendix.

<sup>b</sup>Majority is >50%. Plurality is the physician who provides the largest fraction of care and at least 30%.
Table 2
Assignment of Care Under the 12 Different Attribution Rules Used to Calculate Physician Cost Profile Scores

<table>
<thead>
<tr>
<th>Title of Attribution Rule</th>
<th>Number (%) of Episodes of Care That Could Be Assigned to a Physician&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Number (%) of Physicians Who Were Assigned at least 30 Episodes of Care&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Of all professional costs billed by a physician, % included in that physician’s cost profile Mean % (SD)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Of all professional costs in a physician’s cost profile, % billed by the physician profiled Mean % (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Episode, costs, plurality</td>
<td>3,659,349 (55)</td>
<td>2689 (65)</td>
<td>58 (27)</td>
<td>67 (23)</td>
</tr>
<tr>
<td>Episode, costs, majority</td>
<td>2,901,069 (52)</td>
<td>8401 (61)</td>
<td>51 (27)</td>
<td>73 (24)</td>
</tr>
<tr>
<td>Episode, visits, plurality</td>
<td>2,564,837 (53)</td>
<td>8600 (62)</td>
<td>50 (28)</td>
<td>46 (24)</td>
</tr>
<tr>
<td>Episode, visits, majority</td>
<td>2,871,366 (51)</td>
<td>8415 (61)</td>
<td>46 (27)</td>
<td>49 (25)</td>
</tr>
<tr>
<td>Patient, costs, plurality</td>
<td>2,293,098 (41)</td>
<td>7474 (54)</td>
<td>23 (22)</td>
<td>52 (31)</td>
</tr>
<tr>
<td>Patient, costs, majority</td>
<td>1,140,854 (20)</td>
<td>5567 (40)</td>
<td>13 (17)</td>
<td>55 (37)</td>
</tr>
<tr>
<td>Patient, visits, plurality</td>
<td>2,867,732 (51)</td>
<td>6983 (51)</td>
<td>19 (22)</td>
<td>37 (28)</td>
</tr>
<tr>
<td>Patient, visits, majority</td>
<td>1,651,251 (29)</td>
<td>5411 (39)</td>
<td>12 (16)</td>
<td>39 (32)</td>
</tr>
<tr>
<td>Episode, costs, multiple physicians</td>
<td>3,223,882 (58)</td>
<td>8908 (65)</td>
<td>60 (28)</td>
<td>64 (21)</td>
</tr>
<tr>
<td>Episode, visits, multiple physicians</td>
<td>3,015,484 (54)</td>
<td>8928 (65)</td>
<td>54 (29)</td>
<td>47 (21)</td>
</tr>
<tr>
<td>Patient, costs, multiple physicians</td>
<td>2,349,405 (42)</td>
<td>7841 (57)</td>
<td>22 (20)</td>
<td>53 (30)</td>
</tr>
<tr>
<td>Patient, visits, multiple physicians</td>
<td>3,880,845 (69)</td>
<td>9138 (66)</td>
<td>33 (28)</td>
<td>36 (24)</td>
</tr>
</tbody>
</table>

<sup>a</sup>The total number of episodes was 5,602,652. Percentages are based on this number.

<sup>b</sup>The total number of practicing physicians was 13,761. Percentages are based on this number.

<sup>c</sup>Standard Deviation
Table 3

Percentage of Physicians Assigned a Different Cost Category when Compared to Default Attribution Rule

<table>
<thead>
<tr>
<th>Attribution Rule</th>
<th>Physicians Assigned a Different Cost Category Using this Attribution Rule %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Episode, costs, majority</td>
<td>17</td>
</tr>
<tr>
<td>Episode, visits, plurality</td>
<td>38</td>
</tr>
<tr>
<td>Episode, visits, majority</td>
<td>36</td>
</tr>
<tr>
<td>Patient, costs, plurality</td>
<td>47</td>
</tr>
<tr>
<td>Patient, costs, majority</td>
<td>56</td>
</tr>
<tr>
<td>Patient, visits, plurality</td>
<td>54</td>
</tr>
<tr>
<td>Patient, visits, majority</td>
<td>61</td>
</tr>
<tr>
<td>Episode, costs, multiple physicians</td>
<td>35</td>
</tr>
<tr>
<td>Episode, visits, multiple physicians</td>
<td>46</td>
</tr>
<tr>
<td>Patient, costs, multiple physicians</td>
<td>46</td>
</tr>
<tr>
<td>Patient, visits, multiple physicians</td>
<td>53</td>
</tr>
</tbody>
</table>

*Default rule is Episode, Costs, Plurality. The analyses were limited to the 9,741 physicians who could be assigned at least 30 episodes under any of the 12 rules. For each attribution rule, the results were used to place individual physicians into 1 of 4 categories: low cost, average cost, high cost, or low sample size (> 30 episodes). Disagreement sometimes occurred because the paired rules assigned a physician to 2 different cost categories or because 1 of the rules assigned a physician to a cost category and the other assigned the physician to the low sample size category.*