

CHAPTER

3

**Physician resource use
measurement**

Physician resource use measurement

Chapter summary

Measuring physician resource use and confidentially sharing the results with physicians is one option that might help to address variation in physician practice patterns and Medicare's unsustainable rate of spending growth. In 2005, the Commission recommended that Medicare measure physician resource use and share the analysis results with physicians in a confidential manner. Through the Medicare Improvements for Patients and Providers Act of 2008, the Congress enacted the Commission's recommendation, and CMS has begun a phased implementation of the program, making this an opportune time to detail how the Commission envisions that Medicare's physician resource use measurement program should work.

The Commission has proposed several policy principles to guide Medicare's physician resource use measurement program. These principles include, among others, adopting a methodology for measuring resource use that is transparent to all physicians under observation, ensuring that physicians are able to actively modify their behavior on the basis of the feedback provided, risk adjusting clinical

In this chapter

- Medicare's physician resource use measurement program should follow several policy principles
- Other issues important to physician resource use measurement include stability of results over time and attribution methods
- Conclusion

data to ensure fair comparisons among physicians, and obtaining ongoing feedback from the physician community on CMS's measurement methods and other aspects of the program.

The Commission has also continued to assess its own physician resource use analyses, specifically examining the stability of results over time and studying alternative ways to attribute utilization and costs to physicians. Analyses conducted by the Commission found a high degree of stability in physicians' efficiency scores over time, suggesting that the episode grouper software identifies outlier physicians consistently across years. Our analyses also found that various methods for attributing episodes to physicians have advantages and drawbacks, suggesting that CMS may want to consider more than one attribution method when its physician resource use measurement program is fully implemented. ■

Background

Measuring physician resource use and confidentially sharing the results with physicians is one option that might help to address variation in physician practice patterns and Medicare's unsustainable rate of spending growth. In its March 2005 report, the Commission recommended that Medicare measure physician resource use and share the analysis with physicians in a confidential manner (MedPAC 2005). The Congress enacted this recommendation in the Medicare Improvements for Patients and Providers Act of 2008 (MIPPA) (§131), which requires the Secretary of Health and Human Services to establish a physician feedback program using claims data to provide confidential feedback reports to physicians that measure the resources used to provide care to Medicare beneficiaries. Medicare has already begun work referred to as the Physician Resource Use Measurement and Reporting Program to comply with the MIPPA mandate and to test several characteristics of the program.

Measuring physician resource use is one option to help to address variation in practice patterns and Medicare spending growth

Slowing the increase in Medicare spending is urgent. Medicare's rising costs threaten to place a significant burden on taxpayers. Even the current level of spending may be considered unaffordable as it crowds out other budget priorities and strains financing sources (e.g., the Part A trust fund is now projected to be insolvent in less than 10 years) (Boards of Trustees 2009). Expenditure levels and growth also directly affect beneficiary out-of-pocket costs through higher Part B and supplemental insurance premiums as well as higher cost sharing.

Currently, the government's budgetary mechanism to address rising growth in Medicare expenditures for physician services calls for significant cuts to physician fees. It uses a spending target system, called the sustainable growth rate (SGR) system. The SGR system is designed to offset—through physician fee reductions—spending that exceeds established targets. As designed, the SGR system is inequitable in that all physicians are subject to the consequences (fee cuts) of excess spending that stem from excessive use of resources by only some physicians. In recent years, the Congress has intervened to avoid the fee cuts resulting from aggregate spending on physician services consistently exceeding targets. There is concern that over the long run, sustained payment

reductions, such as those called for by the SGR system, could threaten beneficiary access to services.

While seeking to remedy the SGR problem and make Medicare more sustainable, policymakers have become increasingly interested in examining variation in the use of resources by physicians. Research shows that Medicare spending per beneficiary varies widely across regions of the country, that more variation exists in physicians' practice patterns than can be explained by differences in patients' health status alone, and that areas with more spending do not have better quality outcomes. One study shows dramatic differences in Medicare expenditures among physicians within the same geographic area as well as across areas (GAO 2007). Such variation suggests that an opportunity exists to reduce and redistribute spending to achieve greater efficiency—that is, to get better value—without sacrificing quality. If the physician community were able to glean new insights from analyses comparing physicians' resource use, innovations that improved efficiency—in terms of both quality of outcomes and quantity of resources used—could result.

Measuring physician resource use and confidentially sharing the results with physicians is one option—among several the Commission has discussed—that might help to address Medicare's unsustainable rate of spending growth. In contrast to the inequity of the SGR system, the major advantage of this option is that it would encourage individual accountability among physicians by showing them how their practice patterns affect their patients' total resource use.

Physicians are unique among providers in terms of their ability to drive total resource use. Physicians determine the services they deliver to their patients and influence the care other providers deliver. Under Medicare payment policies, physicians generally receive a separate payment for each individual service they provide.¹ Thus, Medicare spending increases as the volume or intensity of services physicians provide and prescribe increases. In contrast, Medicare pays most other providers a fixed amount for a bundle of services, such as an inpatient hospital stay or a 60-day spell of home health services. This is not to say that bundled payments solve all problems; the Commission has suggested ways to improve payment systems for many types of providers, including those with bundled payments. However, physicians are at one end of the spectrum in terms of fee for individual services, which, as a payment system, presents unique problems.

Providing confidential feedback could alert physicians to inefficient practice patterns they may not be aware of, spurring them to examine and change their practice styles. Providing such feedback directly to physicians has been shown to have a statistically significant, if small, downward effect on resource use (Balas et al. 1996, Schoenbaum and Murray 1992). Because Medicare is the largest single purchaser of health care, its feedback on resource use measurement is likely to be more successful than previous experience in the private sector. The potential success of Medicare's program will depend in part on a significant investment of resources—in terms of dollars and administrative flexibility. In addition, because Medicare's reports would be based on more patients than reports produced by private plans, they may have greater validity and acceptance from physicians.

MIPPA mandate to establish physician feedback program includes program design flexibility

MIPPA (§131) requires the Secretary of Health and Human Services to establish a physician feedback program using claims data to provide confidential feedback reports to physicians measuring the resources used to provide care to Medicare beneficiaries (MIPPA 2008).

It grants the Secretary broad flexibility in designing the program. The Secretary may choose to use data from other sources in addition to claims, provide feedback to individuals and physician groups, and include feedback on both utilization and quality of care. The mandate also permits measuring resources on a per episode basis, a per capita basis, or both. The Secretary may adjust data used for the feedback reports for beneficiaries' health status and other characteristics.

MIPPA also grants the Secretary flexibility to focus the physician feedback program on:

- specialties that account for a significant share of Medicare spending,
- physicians who treat high-cost or high-volume (or both) conditions,
- physicians who use a larger amount of resources than other physicians,
- physicians practicing in certain geographic areas, and
- physicians who treat no fewer than an established minimum number of beneficiaries.

MIPPA required that the Secretary implement the program by January 1, 2009, and conduct education and outreach activities for physicians as part of the feedback program. MIPPA requires the Government Accountability Office to evaluate the physician feedback program by March 1, 2011.

Medicare has begun the physician feedback program

Medicare has begun to test ways to measure physician resource use, distinguish among practice patterns, and share results confidentially with physicians. The work, referred to as the Physician Resource Use Measurement and Reporting Program, complies with the MIPPA physician feedback mandate and will evolve based on experience gained in phases (CMS 2008a). Phase I of the Physician Resource Use Measurement and Reporting Program uses per capita and per episode measurement based on two commercially available software packages (the same ones we have used in our analysis: Symmetry Episode Treatment Groups® (ETGs®), developed by Ingenix, Inc., and the Medical Episode Grouper® (MEG®), by Thomson Reuters) to analyze Medicare claims, produce alternative resource use reports (RURs) for several acute and chronic conditions, provide confidential feedback to selected physicians, and conduct one-on-one interviews with a sample of physicians who receive feedback. The text box provides more detail on episode groupers.

Phase I of CMS's Physician Resource Use Measurement and Reporting Program focuses on four acute conditions (community-acquired pneumonia, urinary tract infection, hip fracture, and cholecystitis) and four chronic conditions (congestive heart failure, chronic obstructive pulmonary disease, prostate cancer, and coronary artery disease with acute myocardial infarction) (CMS 2008b). For these conditions, the program compares physicians with their specific specialty peers and with more general, aggregated peer groups. The program relies on physician-designated specialty, but physicians can have multiple specialties and may treat different conditions, especially across geographic areas. In our own analyses, we discovered—by working backward from a condition rather than by physician-designated specialty—that conditions were largely treated by a few expected specialties but sometimes were treated by unexpected specialties. These rates differed by condition, but instances occurred for which a not insignificant share of physicians treating a condition were of an unexpected specialty (e.g., orthopedic surgeons treating acute myocardial infarction). CMS's program similarly works backward from a condition to create peer groups that cross specialty designations but tend to treat

Episode groupers

Episode groupers are software packages that use clinical logic to assign claims to clinically distinct episodes of care—a series of clinically related health care services over a defined time period, such as all claims related to a patient’s diabetes condition. Episode groupers use all types of health care claims: inpatient admissions, physician visits, other outpatient services, and prescription drugs. They risk adjust by controlling for patients’ comorbid conditions and other characteristics as well as the severity levels of each condition, allowing episode groupers to make more like-to-like comparisons by comparing similar episode–comorbidity–severity combinations rather than comparing all beneficiaries.

A physician’s resource use for selected episodes of care can be compared with the average resource use for similar episodes by peers. This episode-focused comparison may provide more detailed and thus more actionable information than analyses that look at all types of care provided in a physician’s practice. For example, a physician might treat certain patients or conditions in a more resource-intensive manner than others, but when all the physician’s patients are combined in an analysis of per capita spending, the physician’s use of resources appears to be average. An episode grouper has the potential to identify differences in physicians’ practice patterns as well as to examine physicians’ treatment of certain patients or conditions relative to their peers (e.g., excessive use of advanced imaging). ■

similar conditions. CMS’s program is also designed to test multiple geographic areas for comparison (national, state, and hospital service area).

Phase I of CMS’s program tests three risk-adjustment approaches. All three approaches use age, sex, and episode severity. The second approach adds beneficiary overall health status (using hierarchical condition category (HCC) scores). The third approach is similar to the second but adds local area characteristics (county physician supply, average income, and racial and ethnic demographics).

Phase I of the program tests six approaches to attribute episodes to physicians:

- Identify the physician billing most evaluation and management (E&M) visits (plurality).
- Identify the physician billing most E&M visits and accounting for at least a minimum share of total episode costs (plurality-minimum).
- Identify the physician billing most “established patient” E&M visits for chronic conditions only (plurality-established).²
- Attribute the entire cost of an episode to each physician billing for any E&M visit or procedure in the episode (multiple-even).

- Attribute the episode cost to each physician in proportion to billed E&M visits in the episode (multiple-proportional).
- Attribute the entire episode cost to the physician billing for the episode’s first E&M visit for acute episodes only (first contact).

(See attribution discussion on p. 70.)

Phase I also tests several comparison approaches. In principle, to measure physicians’ efficiency, a physician’s resource use for a given episode must be compared with an expected value, often determined by the average resource use of comparable physicians. Under one approach, the program tests using a mean and a median for the expected resource use comparison. As a variant of that approach, the program also explores the right (high cost) and left (low cost) sides of the physician efficiency distribution. CMS is exploring how cut points for defining cost-inefficient and cost-efficient physicians can be set in multiple ways—e.g., two standard deviations from the mean or top or bottom decile.

To gather physician input, CMS distributed RURs to a sample of about 250 physicians in the 12 sites used for the Community Tracking Survey, plus the Baltimore–Washington, DC, area.³ CMS’s contractor to evaluate the physician feedback program, Mathematica Policy

Research, Inc., conducted one-on-one interviews with samples of physicians who received feedback. Physicians were asked their opinions of the alternative RURs and methodologies, especially risk adjustment, attribution, benchmarks, per capita measures, composite measures, details about type of cost or service, and RUR layout.

Informed by the results of phase I of the program, CMS will implement phase II, which may expand the evaluation of physician feedback by including additional specialties, conditions, and geographic areas and including feedback on quality measures.

Significant investment is needed for Medicare’s physician feedback program to evolve from testing various features with a limited number of physicians to a large, widespread program measuring resource use and giving feedback to many physicians. Developing and implementing transparent Medicare-specific measurement methodology, gathering physician input, focusing on outreach and education, and conducting many other resource-intensive activities will shortly be necessary to give the feedback program a chance to achieve its goals. Shortchanging any of these activities risks the viability of the entire physician feedback program. Calling on Medicare to become a value-based purchaser through activities like physician feedback will require a much larger investment in CMS—in terms of both dollars and administrative flexibility.

Medicare’s physician resource use measurement program should follow several policy principles

Given what has recently occurred—we recommended, MIPPA enacted, and CMS has begun a program to implement physician resource use measurement and feedback—it is an opportune time for us to outline how we envision that Medicare’s physician resource use measurement program should work. The measurement and feedback program has the greatest chance of achieving the goals of promoting efficiency and discouraging inefficiency if it follows key policy principles, such as adopting a transparent measurement methodology, reaching out to the physician community for input, focusing on education and outreach, and improving the program over time.

Anticipation of Medicare’s physician resource use measurement program has led to calls to expand the

program from confidential feedback to other activities, such as public reporting, payment adjustments, and aligning efforts with private payers. These kinds of activities would be transformative steps toward Medicare becoming a value-based purchaser. However, the Commission is concerned that expanding beyond confidential feedback too rapidly could lead to a flawed physician resource use measurement program and that even the appearance of moving too rapidly could undermine physician and beneficiary confidence in the program. The Commission has recommended that Medicare design physician resource use measurement so as to be prepared for any eventual public reporting and payment adjustments. First, Medicare and the physician community will need time to learn from the experience of confidential feedback. In the meantime, Medicare is gaining related experience through public reporting of physicians participating in the Physician Quality Reporting Initiative, paying physicians based on quality reporting, and sharing claims data with other entities through the Generating Medicare Performance Results Project and Chartered Value Exchanges (Milgate 2008). Together, these experiences should inform decisions about the future direction for the physician resource use measurement program.

Adopt measurement methodology that is transparent

Before Medicare finalizes the confidential physician feedback program, it should make publicly available an explanation of its measurement methodology and a description of the data sources used. Currently, CMS’s Physician Resource Use Measurement and Reporting Program relies on commercially available episode grouper software packages, which allows the agency to evaluate the software packages’ features that can be included in a Medicare-specific, open source software package. However, the Commission has never expected the Medicare program to purchase off-the-shelf software with proprietary, black-box methodology; CMS does not normally pursue such a strategy. It usually contracts with vendors to develop tailored programs, such as is done for the Medicare severity–diagnosis related groups used in the inpatient hospital payment system. Similarly, the episode grouper CMS finally decides on for its physician resource use measurement program should use a Medicare-specific, transparent method. (Existing episode grouper methodology has become more transparent. In March 2009, Ingenix, Inc., released its ETG measurement methodology for public review and comment (Ingenix

2009).) The program CMS selects could be provided by one of the existing episode grouper software companies, tailored to suit Medicare's needs. Existing episode grouper software has been used by private payers. Since Medicare was not a customer until recently, the software was not developed with the program's unique characteristics in mind. Therefore, existing software may need to be modified to suit the Medicare program.

Adopt measurement refinements as program evolves

Medicare should not wait until the measurement methodology is perfected to implement the physician feedback program. Since the proposed resource use measurement program relies on confidential feedback to educate physicians, it should begin as soon as possible with as many measures as are ready. Measures can be refined and new ones added over time.

Ideally, changes in physicians' year-to-year resource use measurement results should be due to changes in their practice patterns alone rather than to changes in measurement methods. However, this program will be a new endeavor for Medicare. It is unrealistic to expect the measurement methodology used in the first year to remain unchanged in future years. One way to help deal with these changes is to pilot test any future refinement by including new measures, highlighted as such, in detailed feedback for a year or two before including them in overall scores.

Ensure that feedback is actionable

Feedback should include detailed breakouts—such as by type of service, provider, and condition—in addition to overall scores in such a way that it is clear to physicians which aspects of their practice patterns they should act on (Figure 3-1, p. 68). (This sample feedback form has not been used in any feedback programs and is provided for illustrative purposes. Other feedback forms, including the one used by the Medicare program, differ from this one.) For example, some physicians treat diabetic patients in a more resource-intensive manner or use more intensive imaging services than their peers. Providing detailed information in addition to aggregate measures makes physician feedback more actionable by identifying differences in practice patterns that influence physicians' overall feedback results.

Medicare will need to balance the value of providing physicians with detailed information about total resource use with beneficiaries' right to privacy.

Detailed information about services provided by other physicians and other types of providers would greatly enhance physicians' ability to evaluate their practice and referral patterns. Without this type of information, it is unreasonable to expect physicians to significantly improve care coordination and chronic disease management and achieve many similar policy goals. However, sharing information on care not directly provided by that physician—even providing a list of other physicians caring for a patient—can easily run afoul of beneficiary (and other physician) privacy rights. Perhaps a way for the Medicare program to balance these two competing goals is to ensure that sufficient information is provided to physicians to make them aware of their relative performance and to strive to provide as much additional information as possible to support care coordination. In other words, information about resource use can be aggregated in a way that physicians can see that they are more or less efficient than their peers without disclosing information about specific services provided to individual patients or by individual providers. For physician feedback to also support care coordination, the Medicare program should strive to include more detailed information, especially the names of other providers involved in the physician's episodes. This information would not reveal the average efficiency scores of other physicians, but it would allow physicians to act on feedback by discussing treatment patterns with their colleagues. The ability to call other physicians to discuss the treatment of patients is at the heart of care coordination.

We want to be clear that the feedback will not answer all questions about how to improve practice patterns for greater efficiency. The success or failure of the feedback program will depend on Medicare's ability to forge a collaborative partnership with physicians and on the physician community's willingness to embrace thoughtful examination of their practice patterns. Physicians will have to come together in professional societies and other organizations to learn from feedback and discuss how best to improve efficiency and then act on these decisions.

Risk adjust data to ensure appropriate comparisons

MIPPA gives the Secretary discretion to adjust data used for the feedback reports for beneficiaries' health status and other characteristics. The program must make such adjustments to measure resource use as accurately as possible.⁴ Risk adjustment can help to indicate when resources are overprovided to healthy patients as well as when resources are underprovided to patients in greater need.

FIGURE 3-1

Example of the first page of a physician feedback form

Summary of patterns of care

Name:	John Smith, MD	Peers' average
Peer group:	Cardiology	Cardiology
Total spending:	\$XXX,XXX	\$YYY,YYY
Number of patients:	XX	YY
Cost per patient:	\$X,XXX	\$Y,YYY
Number of episodes:	XX	YY
Cost per episode:	\$X,XXX	\$Y,YYY
Episodes per patient:	X	Y
Average patient health status:	X.XX	Y.YY

Summary of top 10 episodes by total cost

Episode	Number of episodes	Your cost per episode	Peers' cost per episode	Your cost index
Total (all episodes)	XX	\$X,XXX	\$Y,YYY	1.2
<i>Your type of service cost index relative to peers'</i>				
Hypertension	XX	\$X,XXX	\$Y,YYY	1.3
<i>Your type of service cost index relative to peers'</i>				
Coronary artery disease	XX	\$X,XXX	\$Y,YYY	1.0
<i>Your type of service cost index relative to peers'</i>				
Arrhythmias	XX	\$X,XXX	\$Y,YYY	0.8
<i>Your type of service cost index relative to peers'</i>				

Note: E&M (evaluation and management), PAC (post-acute care).

Existing episode grouper software packages do risk adjust for disease stages, patients' comorbid conditions, and other characteristics. The methods are somewhat similar to those included in the CMS–HCC risk-adjustment method used by Medicare to adjust Medicare Advantage plans' payments for the health status of their enrollees. An appropriate risk-adjustment method, such as CMS–HCC, should be used for beneficiary-level measures of resource use, such as per capita utilization.

We have conducted our analysis by comparing physicians only with others in the same specialty and the same metropolitan statistical area (MSA) (see text box, p. 71). This is a conservative methodology, in that it is less likely to find differences among physicians' efficiency than comparing against national averages. It also helps to offset one of the limitations of risk adjustment based on diagnoses: Individuals who receive more health services are likely to have more (and more serious) diagnoses coded than individuals who receive fewer health services, even when factors other than health, such as the supply of specialists, influence the amount of health services.

Some researchers have suggested comparing across specialties or geographic areas. If Medicare were to do so, it would be critical that the program not adjust away any spending differences that Medicare should be concerned about, such as spending differences correlated with differences in the supply of specialists. In other words, risk adjustment is designed to help match spending to patients. One expects patients who are old and sick to cost more than those who are young and healthy. Physicians should not be held accountable for these resulting spending differences; therefore, we try to risk adjust for these differences. However, patients' costs are influenced by other factors, such as the types of physicians they visit and where they live. Physicians should be held accountable for spending driven by some of these factors, and therefore one should measure these spending differences and not risk adjust for them.

Use multiple measures of resource use to produce more meaningful results

The physician feedback program should have the flexibility to measure physician resource use on both a per episode and a per capita basis (Figure 3-1). Both methods analyze claims data to better understand physicians' practice patterns. Episode-based methods group claims into clinically distinct episodes of care and then compare resource use for similar episodes. Per capita–based methods analyze total resource use for each patient or each

beneficiary in an area. Together these measures more fully capture the relevant characteristics of physicians' practice patterns by revealing physicians' resources used in an episode and the number of episodes per patient. Relying on either measure alone could mask differences between physicians and even allow gaming such as generating more episodes to appear more efficient on a per episode basis. Additional measures—such as rate of prescribing generic drugs and use of basic versus advanced imaging—should also be included when warranted to produce a more complete picture of resource use. As a practical matter, however, the program cannot wait for implementation until all these measures are ready. Instead, the program should begin with as many appropriate measures as it reasonably can and transition to implementation of the full measurement set. The program should be flexible enough to weight or even exclude measures where appropriate.

Obtain physician input on resource use measurement program

The program will need to balance Medicare's need to make methodology decisions necessary to begin implementation with physicians' right to be fairly measured. In seeking this balance, the program will need—and has already begun—to obtain physician input. First, CMS's Physician Resource Use Measurement and Reporting Program obtains physician input through one-on-one interviews with select physicians who receive feedback under the program. Second, the Medicare program will need to continue to obtain physician input over time. To this end, the agency will need to consult with physicians and may want to consider working with formal physician advisory boards and through informal interactions with physician organizations and individual physicians. Third, CMS should include, as part of the physician proposed rule published each year, a description of planned changes to the resource measurement program's methodology, feedback process, or other issues. Finally, once Medicare implements confidential feedback, as long as it seeks physicians' reactions, it will essentially operate a continuous physician comment period.

Provide feedback to nonoutlier as well as outlier physicians

In principle, the feedback component of the resource use measurement program is intended to change the behavior of physicians. Some suggest providing feedback only to physicians whose clinical practices fall outside the norm (outliers), creating system inefficiency, adversely affecting quality, or both. In this case, feedback would

**TABLE
3-1****Physicians' 2002 efficiency scores are highly correlated with their 2003 scores, using either multilevel or Monte Carlo models**

MSA	Multilevel	Monte Carlo
Boston, MA	0.90	0.87
Greenville, SC	0.91	0.89
Miami, FL	0.88	0.86
Minneapolis, MN	0.86	0.84
Orange County, CA	0.89	0.84
Phoenix, AZ	0.90	0.88
Total	0.89	0.87

Note: MSA (metropolitan statistical area). Physicians with fewer than 20 episodes were excluded from the analysis. Correlation coefficients measure how the ranks of items in two different lists compare. A perfect correlation of 1.00 means that the items are at exactly the same rank in both lists. A coefficient of 0 means that there is no relationship between the rank of items on the two lists.

Source: MedPAC analysis of 100 percent sample of 2001–2004 Medicare claims using the Thomson Reuters Medical Episode Grouper®.

be provided only to physicians whose resource use exceeded a certain threshold, physicians who treated higher cost or more common conditions, or other subsets of physicians. Focusing on such outliers would be more feasible administratively and less costly, while offering the opportunity for some positive impact by altering the practice patterns of the most inefficient physicians.

Alternatively, advantages exist to providing feedback to most Medicare physicians. Giving detailed feedback to physicians across the entire efficiency distribution would allow even nonoutliers to recognize any of their own inefficient practices—such as ordering duplicative tests or overusing advanced imaging—and to work toward improving them. As a practical matter, however, the program could not be expected to provide feedback to all physicians who treat Medicare beneficiaries. Instead, the program is designed to measure and compare physicians' resource use with their peers' use only if they provide enough of a beneficiary's care to be considered responsible for the beneficiary or a given episode of care and if they treat enough beneficiaries and episodes to warrant comparison.

Measure and provide feedback to both individual physicians and group practices

The physician feedback program should use individual physicians as the basic building block of resource

use measurement methodology but be capable of aggregating these measures in multiple ways—such as by physician group practice or by accountable care entities—for confidential feedback. This capacity will allow the program maximum flexibility in applying the measurement results in multiple ways to tailor feedback reports to best suit physicians' preferences. It also will allow the program to measure the nearly 40 percent of physicians who continue to work as solo practitioners (Hing and Burt 2008).

Focus on education and outreach

MIPPA requires that Medicare conduct education and outreach activities as part of the physician feedback program. Merely mailing physicians a feedback report is not enough. At a minimum, physicians need to be able to contact someone for answers to their questions. We learned through site visits that education and outreach are often neglected aspects of physician resource use measurement programs and that this oversight impairs these programs' chances of success. Given CMS's limited resources and numerous responsibilities, these new efforts will be challenging. CMS could partner with other entities, including physician organizations and specialty societies, to support physicians in interpreting feedback reports and using them to improve practice patterns. Another possible approach is to redirect the Quality Improvement Organizations' scopes of work to these efforts.

Other issues important to physician resource use measurement include stability of results over time and attribution methods

In our ongoing physician resource measurement analyses of using Medicare claims and episode grouper software, we most recently explored the stability of results over time and the trade-offs among different methods for deciding which physicians to hold responsible for a beneficiary's episode of care. The strong correlations in physicians' efficiency scores over time suggest that those scores are generally stable over time. The existence of advantages and drawbacks of various attribution methods means that CMS may need to consider using more than one attribution method in its fully implemented physician resource use measurement program.

Statistical methodology

The analysis to evaluate the year-to-year stability of physicians' efficiency scores was conducted by Thomson Reuters using the firm's Medical Episode Grouper[®] (Thomson Reuters 2009b). We used two statistical models to compare physicians' observed resource use with their peers' (expected) resource use. The two models build on the simple observed-to-expected ratios that are generally used. (Peers are defined as physicians in the same specialty in the same metropolitan statistical area (MSA).) In each case, the observed resource use is the same; what differs is the measure of expected resource use. Both the multilevel regression and the Monte Carlo randomization models calculate expected resource use taking into account case-mix variation. As in any calculation of a threshold, it is up to the judgment of the analyst to decide what

threshold defines an outlier. To be conservative, we chose to set relatively high thresholds for identifying outlier physicians.

We used Medicare claims for beneficiaries living in six MSAs: Boston, MA; Greenville, SC; Miami, FL; Minneapolis, MN; Orange County, CA; and Phoenix, AZ. We standardized payments by excluding variation in resource costs due to geographic differences in input costs or policy considerations (e.g., teaching hospital payments). For this analysis, we wanted resources spent on, for example, a hospital admission for stroke to be comparable across geographic areas and facility types. Removing the effects of payment policies allowed us to conclude that underlying differences in clinical resource use were due to differences in practice patterns. ■

Physicians' efficiency scores are stable over time

To determine the stability of physician efficiency scores developed from our resource use analyses, we compared physicians' efficiency scores (measures of relative resource use) over two points in time—2002 and 2003 (Thomson Reuters 2009b). Correlations between 2002 and 2003 efficiency scores, weighted by each physician's average number of episodes per year, are shown in Table 3-1. The correlations were high, indicating good year-to-year stability in the efficiency scores using two statistical methods—one based on a multilevel model and the other based on a Monte Carlo model (see text box on statistical methodology). Physicians with high efficiency scores in 2002 tended to have high scores in 2003 and vice versa.

In addition to comparing all physicians' efficiency scores year to year, we further analyzed physicians whose efficiency scores qualified them as outliers in 2002. A physician was considered an outlier in 2002 if the physician's observed score differed statistically from his or her risk-adjusted expected score at the 0.0001 significance level. Further, that physician would be considered an outlier in both 2002 and 2003 if the physician's 2003 observed score also differed from his or her risk-adjusted score at the 0.05 level of statistical significance. Setting the threshold in this way gives us great confidence that

we are not identifying false-positive outliers—that is, physicians whose practice styles appear unusually high because of random fluctuation. Using this definition, we found that of the 611 outliers in 2002, 572 (94 percent) were also outliers in 2003. The 6 percent of physicians who were labeled outliers in 2002 but not in 2003 may have been “false positives.” Alternatively, it is possible that these physicians were truly outliers in 2002 and truly not outliers in 2003. One would expect some natural variation in physicians' efficiency from year to year.

Trade-offs between using single and multiple attribution

One of the main goals of grouping claims into episodes is to attribute the care provided during those episodes to particular physicians and ultimately to quantify how efficient their use of resources was for their patients. In the private sector, some plans—such as HMOs that use gatekeepers—formally assign patients to a primary care physician, so attribution is relatively straightforward. However, in other plan types and in the Medicare fee-for-service program, patients have the freedom to see any physician. This structure makes attribution less straightforward. In these cases, users of episode grouper software rely on patterns in claims data to attribute episodes to physicians.

**TABLE
3-2**

The number of physicians, the number of physicians who submitted any claim for at least 20 episodes, and the mean number of episodes per physician varied by MSA

Physicians who submitted any claim for at least 20 episodes

MSA	Total physicians		2002			2003		
	2002	2003	Number	Percent	Mean episodes per physician	Number	Percent	Mean episodes per physician
Boston, MA	16,495	17,191	11,111	67%	314	11,615	68%	337
Greenville, SC	2,715	2,948	2,137	79	623	2,254	76	613
Miami, FL	6,331	6,654	4,787	76	409	4,969	75	417
Minneapolis, MN	10,015	10,565	7,098	70	271	7,486	71	268
Orange County, CA	6,570	6,835	4,450	68	343	4,715	69	347
Phoenix, AZ	8,338	8,946	5,950	71	328	6,411	72	338
Total	50,464	53,139	35,533	70	343	37,450	70	352

Note: MSA (metropolitan statistical area). The 20-episode minimum was selected for illustrative purposes only.

Source: MedPAC analysis of 100 percent sample of 2001–2004 Medicare claims using the Thomson Reuters Medical Episode Groupers®.

A key question about how to attribute episodes to physicians is whether to use single attribution (holding a single physician responsible for the care provided) or multiple attribution (holding more than one physician responsible for the care provided). Single attribution is designed to identify the “decision maker,” perhaps the primary care physician, and hold this individual responsible for all care rendered. Multiple attribution acknowledges that the decision maker, if there is one, has incomplete control over treatment by specialists and other physicians, even if the decision maker referred the patient to those other physicians.

For our analysis to date we have used a single attribution method with a 35 percent threshold of E&M dollars; that is, if a physician was responsible for at least 35 percent of the E&M dollars in a given episode, we attributed that episode, and all its costs, to that physician. Policymakers should not interpret our use of a 35 percent threshold of E&M dollars as a recommendation. In fact, attribution methods and their policy implications warrant further discussion.

Results of attribution analysis

There are significant trade-offs between attribution methods, so we wanted to examine whether a quantitative analysis yielded a clearly preferable method. To explore attribution methods, we compared resource measurement

results using multiple attribution with those using single attribution. Table 3-2 shows—for each of the 6 MSAs examined in 2002 and 2003—the total number of physicians, the number of physicians who submitted any claim for at least 20 episodes (the required minimum for a physician to be included in the analysis), and the average number of episodes per physician (Thomson Reuters 2009a). (Researchers who use episode groupers generally agree that it is statistically invalid and unfair to calculate efficiency scores for physicians with too few episodes. We selected 20 episodes as a minimum for illustrative purposes; this selection should not be viewed as a policy recommendation.) The number of physicians who submitted any claim for at least 20 episodes was generally greater than the number of physicians who were attributed responsibility for episodes, which varied by attribution method. Within each MSA, the three sets of numbers were similar for the two years, whereas the three sets of numbers varied substantially by MSA. For example, the total number of physicians in each MSA ranged from fewer than 3,000 in Greenville to more than 16,000 in Boston.

We examined four methods of attributing episodes to physicians, based on a combination of the following variables: episodes associated with a single physician, episodes associated with multiple physicians, physician expenditures identified for E&M services (E&M dollars), and physician expenditures identified for all Medicare-

covered services (total dollars).⁵ The four attribution methods are depicted in Figure 3-2. All four methods calculate ratios of observed-to-expected (O/E) resource use, using an average of episode-level O/E ratios, which we found to be preferable to calculating O/E ratios using ratios of average dollars.⁶

We compared results for each of the attribution methods. Both the choice of using single versus multiple attribution and using E&M versus total dollars affected the share of physicians to whom at least 20 episodes are attributed. As one would expect, the number of physicians who are attributed at least 20 episodes differed by attribution method, with multiple attribution methods resulting in more physicians meeting this threshold than single attribution methods (Table 3-3). Similarly, attribution methods based on total dollars resulted in more physicians meeting the 20-episode threshold than those based on E&M dollars, but this difference (total versus E&M) was smaller than the difference between multiple and single attribution methods.

TABLE 3-3

Multiple attribution based on total dollars produced the greatest number of physicians meeting the 20-episode minimum requirement for inclusion in our measurement analysis

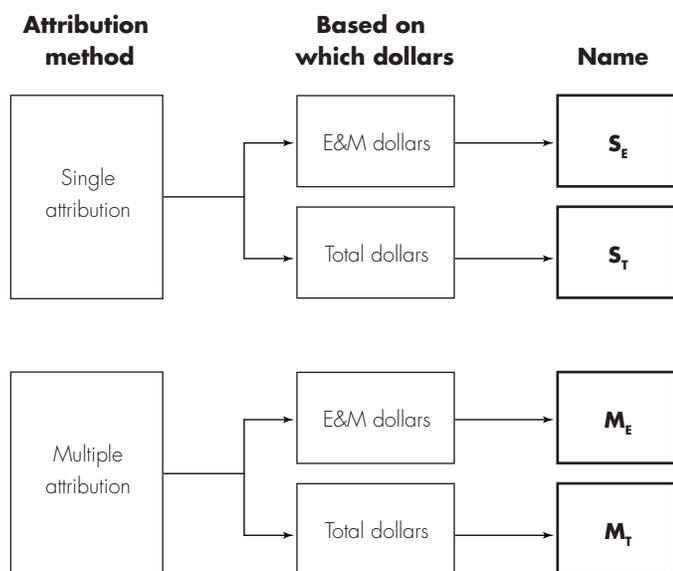
Attribution method	Name	Percentage of physicians attributed at least 20 episodes
Multiple attribution		
Based on total dollars	M_T	70.4%
Based on E&M dollars	M_E	55.6
Single attribution		
Based on total dollars	S_T	53.9
Based on E&M dollars	S_E	48.0

Note: E&M (evaluation and management). The 20-episode minimum was selected for illustrative purposes only.

Source: MedPAC analysis of 100 percent sample of 2001–2004 Medicare claims using the Thomson Reuters Medical Episode Grouper®.

FIGURE 3-2

Four methods for attributing episodes to physicians



Note: E&M (evaluation and management). Single attribution methods result in a weight of 1 for the physician attributed responsibility for the episode and 0 for all other physicians involved with the episode.

We were also interested in the effect different attribution methods had on physicians' O/E ratios. The correlations among the four attribution methods are shown in Table 3-4 (p. 74). These correlations are based on physicians to whom at least 20 episodes were attributed under both methods being compared, and the physicians were weighted by the average number of episodes in both methods. For example, if a physician had 20 episodes attributed using the multiple attribution based on E&M dollars (M_E attribution method) and 30 episodes using the single attribution based on E&M dollars (S_E attribution method), that physician would have been included in the calculation of the correlation between M_E and S_E with a weight of 25. Only the correlations for 2003 are shown; there was little difference between the 2002 correlations and the 2003 correlations.

Single attribution and multiple attribution indices give very similar results based on E&M dollars (0.97); the correlation is similarly high for indices based on total dollars (0.95). The correlations are somewhat lower—between 0.86 and 0.91—when comparing S_E and M_E indices with indices based on total dollars (S_T and M_T). Therefore, the attribution method selected does not significantly affect physicians' O/E ratios. Physicians who appear to be efficient (or inefficient) under one attribution

TABLE 3-4

Number of physicians being compared and correlations among attribution methods, 2003

Attribution method	M _E	M _T	S _E	S _T
M _E		29,563	25,529	25,690
M _T	0.91		25,529	25,690
S _E	0.97	0.87		25,690
S _T	0.86	0.95	0.87	

Note: M_E (multiple attribution based on evaluation and management (E&M) dollars), M_T (multiple attribution based on total dollars), S_E (single attribution based on E&M dollars), and S_T (single attribution based on total dollars). Correlation coefficients measure how the ranks of items in two different lists compare. A perfect correlation of 1.00 means that the items are at exactly the same rank in both lists. A coefficient of 0 means that there is no relationship between the rank of items on the two lists. Correlations among attribution methods are shown below the diagonal line. The numbers of physicians attributed at least 20 episodes under both methods being compared are shown above the diagonal line. The 20-episode minimum was selected for illustrative purposes only.

Source: MedPAC analysis of 100 percent sample of 2001–2004 Medicare claims using the Thomson Reuters Medical Episode Grouper®.

method generally appear to be efficient (or inefficient) under other attribution methods.

Finally, we compare the year-to-year stability in physicians’ O/E ratios for the various attribution methods. The year-to-year correlations, shown in Table 3-5, tend to be fairly high for all of the attribution methods. The lowest correlation was 87 percent for the S_E attribution method and the highest correlation was 91 percent for the S_T and M_T methods. The lack of significant effect of attribution method on the year-to-year stability of physicians’ O/E ratios also appears to rule out stability as a factor to use in deciding which attribution method would be optimal.

The lack of clear differentiation among attribution methods based on our statistical analysis means that there is no single “right” answer to the question of how

to attribute episodes to physicians. Therefore, the choice among attribution methods probably comes down to a qualitative decision based on the policy goals of the program. For example, if Medicare would like physicians to focus more on the effects of their referrals, they might select a single attribution method. Alternatively, if Medicare wanted to trigger conversations among physicians caring for the same patient, the program might select a multiple attribution method. The final program may have reason and room for more than one attribution method.

Conclusion

The Commission has recommended that Medicare develop a physician resource use measurement program and confidential feedback program; this program was enacted by MIPPA and is being implemented by CMS. The

TABLE 3-5

Year-to-year correlations of physicians’ observed-to-expected ratios are high for all four attribution methods

Attribution method	Name	Correlation in physicians’ observed-to-expected ratios, 2002–2003
Multiple attribution		
Based on E&M dollars	M _E	0.89
Based on total dollars	M _T	0.91
Single attribution		
Based on E&M dollars*	S _E	0.87
Based on total dollars**	S _T	0.91

Note: E&M (evaluation and management). Correlation coefficients measure how the ranks of items in two different lists compare. A perfect correlation of 1.00 means that the items are at exactly the same rank in both lists. A coefficient of 0 means that there is no relationship between the rank of items on the two lists. Correlations are based on physicians with at least 20 attributed episodes in both years, and each physician was weighted for his or her average number of episodes per year. The 20-episode minimum was selected for illustrative purposes only.
 * For single attribution, the physician with the most E&M dollars is given a weight of 1 for the episode and all other physicians are given a weight of 0.
 ** For single attribution, the physician with the most total dollars is given a weight of 1 for the episode and all other physicians are given a weight of 0.

Source: MedPAC analysis of 100 percent sample of 2001–2004 Medicare claims using the Thomson Reuters Medical Episode Grouper®.

program should begin with the best methodology possible, but it should not be delayed until all methodologic questions are addressed. The measures should be added to and refined as Medicare gains experience. As Medicare and physicians learn from confidential feedback and related activities, this experience should inform decisions about the future direction for the program, such as the possibility of adding public reporting and aligning activities with private payers.

Physicians' efficiency scores are generally stable over time. Attribution is one of the methodology questions that should continue to be explored. The choice between single and multiple attribution cannot be made based on statistical results alone. Therefore, the choice may hinge on other policy considerations, such as how best to spur discussion among physicians about their individual contributions to overall resource use. ■

Endnotes

- 1 There are exceptions—for example, the global surgical bundle pays for physician services before, during, and after operations.
- 2 E&M visits are separated into those for new patients and those for established patients.
- 3 The 12 sites used for the Community Tracking Survey are Boston, MA; Cleveland, OH; Greenville, SC; Indianapolis, IN; Lansing, MI; Little Rock, AR; Miami, FL; Northern New Jersey; Orange County, CA; Phoenix, AZ; Seattle, WA; and Syracuse, NY.
- 4 No risk-adjustment method predicts all costs. There is truly random variation that cannot be predicted at the individual level (e.g., being struck by a bus). Nor should a risk-adjustment method be expected to adjust away all cost differences. There is variation that should be examined by researchers and policymakers (e.g., geographic differences in utilization identified in the Dartmouth Atlas (Wennberg et al. 2008)). Both per capita and per episode methods for risk adjustment have improved over time and will continue to improve in their ability to appropriately account for cost variation. The Commission regularly analyzes potential refinements to risk adjustment. It is important that CMS use the best risk-adjustment methods available and implement refinements over time.
- 5 We examined eight attribution methods and include the four best here. For information on the other four methods, see the multiple attribution report (Thomson Reuters 2009a).
- 6 To calculate a ratio of averages for a given physician, one would calculate the mean of his or her observed payments and then divide this value by the mean of corresponding expected payments. Mathematically, the result differs from calculating an average of episode-level ratios by calculating an O/E ratio for each individual episode and then taking a mean.

References

Balas, E., S. Borch, G. Brown, et al. 1996. Effect of physician profiling on utilization: Meta-analysis of randomized clinical trials. *Journal of Internal Medicine* 11, no. 10: 584–590.

Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. 2009. *2009 annual report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Insurance Trust Funds*. Washington, DC: Boards of Trustees.

Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2008a. Medicare program; payment policies under the physician fee schedule and other revisions to Part B for CY 2009; e-prescribing exemption for computer-generated facsimile transmissions; and payment for certain durable medical equipment, prosthetics, orthotics, and supplies (DMEPOS). Final rule. *Federal Register* 73, no. 224 (November 19): 69775–69824.

Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2008b. VBP initiatives: Physician resource use. Baltimore: CMS. http://www.acc.org/advocacy/advoc_issues/MDI_conf_2008/CMS_PRU%20ACC%202008-10-30_CMS.ppt.

Government Accountability Office. 2007. *Medicare: Focus on physician practice patterns can lead to greater program efficiency*. GAO–07–307. Washington, DC: GAO.

Hing, E, and C. W. Burt. 2008. Characteristics of office-based physicians and their practices: United States, 2005–2006. *Vital and Health Statistics* 13, no. 166. http://www.cdc.gov/nchs/data/series/sr_13/sr13_166.pdf.

Ingenix. 2009. Ingenix makes market-leading Episode Treatment Groups[®] methodology available for public review. March 10. <http://www.ingenix.com/News/Article/92/>.

Medicare Improvements for Patients and Providers Act, Pub. L. No. 110–275, 122 Stat. 2494. 2008. Codified at 42 U.S.C. 1305 note § 131.

Medicare Payment Advisory Commission. 2005. *Report to the Congress: Medicare payment policy*. Washington, DC: MedPAC.

Milgate, K. 2008. *Use of Medicare information to improve health care delivery*. Baltimore: CMS. <http://www.ftc.gov/bc/workshops/hcbio/docs/hcq/kmilgate.pdf>.

Schoenbaum, S., and K. Murray. 1992. *Impact of profiles on Medicare practice*. Conference on profiling no. 92–2. Physician Payment Review Commission, paper no. 4: 72–102.

Thomson Reuters. 2009a. *Multiple attribution of episodes for physician profiling in Medicare*. New York: Thomson Reuters.

Thomson Reuters. 2009b. *The use of an episode grouper for physician profiling in Medicare*. New York: Thomson Reuters.

Wennberg, J., E. Fisher, D. Goodman, and J. Skinner. 2008. *Dartmouth atlas of health care 2008: Tracking the care of patients with severe chronic illness*. http://www.dartmouthatlas.org/atlas/2008_Chronic_Care_Atlas.pdf.

