

Working PAPER

BY GREG PETERSON AND ERIC SCHONE

Rewarding Physicians for Their Patients' Health Outcomes: What Can Medicare Learn from Education's Value-Added Models?

June 2012

ACKNOWLEDGMENTS

The authors would like to thank Tim Lake and Steve Glazerman, researchers at Mathematica Policy Research, for their constructive reviews of earlier drafts of this paper. We would also like to thank several other researchers at Mathematica for their valuable input, including Jeff Ballou, Duncan Chaplin, Arnold Chen, Tom Croghan, Eric Isenberg, Mary Laschober, Myles Maxfield, and Bob Schmitz. Finally, we would like to thank Amanda Bernhardt and Daryl Hall for editing the paper.

ABSTRACT

Starting in 2015, the Centers for Medicare & Medicaid Services (CMS) will begin adjusting payments sent to physicians for Medicare services based on the value of the care provided. The move is one of many efforts in the private and public sectors to promote higher quality, lower cost health care through value-based purchasing. “Value” is defined as a combination of the quality and cost of the care. CMS’s goal is to create incentives for physicians to improve the value of their care. Although the agency would like to incorporate patient outcomes in its assessments of physician quality, this approach raises a key methodological challenge: how does one isolate a physician’s influence on patient outcomes from the many other factors that influence these outcomes, such as a patient’s diet or illness severity?

Many articles in the health literature have examined these issues. The contribution of this paper to the literature is two-fold. It examines:

- How “value-added” models used in education research address a similar challenge when measuring and rewarding teacher performance based on student test scores
- Whether these approaches are applicable to value-based purchasing of health care in general, and of Medicare services in particular

We argue that value-added models are potentially useful in Medicare because they could provide a powerful way to control for patient- and system-level characteristics that are outside a physician’s control but that influence patient outcomes. However, using these models to measure physician performance poses several practical and conceptual challenges:

- Accounting for multi-year relationships between physicians and their patients, as opposed to the typical one-year relationship between teachers and students
- Statistical uncertainty in the value-added estimates, which is also a concern in value-added estimates for teachers
- Limitations in the availability of outcome data.

We discuss these challenges as well as possible strategies for addressing them.

I. POLICY CONTEXT

As of 2015, the 2010 Affordable Care Act directs CMS to adjust payments sent to physicians for Medicare services based on the value of the care provided. “Value” is defined as a combination of the quality and cost of the care. Physicians participating in third-party fee-for-service Medicare are paid separately for each service provided to patients. This approach does not provide strong incentives to physicians to avoid unnecessary services; to track patients over time to ensure that they receive appropriate, evidence-based care; to coordinate with other physicians; or to intervene early to prevent hospitalizations or emergency room visits, which can generate larger expenses for Medicare. Medicare’s “value-based modifier” is intended to provide new incentives for physicians to focus on value by adjusting their fee-for-service payments based on the measured value of their care. For example, if a physician’s care is deemed to be high value, Medicare may pay two percent more for services that the physician provides. The modifier is part of a broader CMS initiative to purchase care in a way that achieves the three-part aim of higher-quality care, better population health, and lower costs. The Affordable Care Act also directs CMS to develop value-based purchasing programs for hospitals, home health care, skilled nursing facilities, and ambulatory surgical centers.

CMS would like to include patient health outcomes in its measure of the quality of care that a physician provides. For example, in recent regulations describing their plans for the modifier, CMS indicated that they plan to include specific clinical outcomes, such as blood-pressure control for patients with diabetes, in their measures of physician performance (Office of the Federal Register 2011a). CMS is also considering improved patient functional status and avoidance of adverse events, such as hospitalizations that could be prevented by good ambulatory care, as outcome measures.

However, in measuring performance based on outcomes—rather than on specific clinical process measures that are more under a provider’s direct control—CMS faces the key methodological challenge of isolating a physician’s influence on the outcomes from the many other factors that also affect health outcomes, such as diet and illness severity. Traditional risk-adjustment methods based on both demographics and diagnoses listed on claims submitted by providers to Medicare for reimbursement do a poor job of capturing the influence of these factors. If this issue is not addressed, Medicare’s payment modifications could be unfair to physicians, creating unintended consequences; for example, physicians might drop from their panel patients who are very sick because they lower the physicians’ measured performance. Indeed, many of those who responded to CMS’s plans for the value-based modifier said that CMS should not hold physicians accountable for factors outside their control and that a move in this direction could discourage physicians from caring for patients with highly complex medical needs (Office of the Federal Register 2011b).

Although many articles in the health literature have addressed this issue, CMS is actively seeking advice on how to properly measure physician performance based on health outcomes, including approaches to isolating the physician’s influence on these outcomes. The contribution of this article is to examine the education literature to determine whether techniques similar to those used in recent educational reforms to measure teacher effectiveness based on student outcomes (test scores) could be applied to Medicare as CMS designs the value-based modifier.

II. THE VALUE-ADDED APPROACH TO ISOLATING TEACHER EFFECTS ON STUDENT OUTCOMES

Over the past 20 years, education researchers have developed methods to isolate a teacher's influence on student achievement (as reflected in test scores) from the many other factors that influence learning. These factors include poverty, peer pressures, a student's aptitude, parental involvement, other teachers, and school policies and resources. Many school districts have begun using the results of these "value-added" models to evaluate their teachers' effectiveness and, in combination with principal and/or peer reviews, determine which teachers earn bonuses, promotions, or tenure.

Despite the many variations in value-added models, they all use longitudinal student test-score data to disentangle the effects of individual teachers on student learning from other influences (McCaffrey et al. 2003). Value-added models frequently use students' past test scores as well as measurable student, peer, and school characteristics to predict the gains in test scores that each student would have from one year to the next, assuming that he or she was taught by an average teacher. The models then compare the actual performance of a teacher's students to these predicted values. A teacher receives a positive value-added score if his or her students, on average, performed better than the predicted values (and a negative score if the opposite is true).

A basic value-added model to estimate the effectiveness of an eighth-grade math teacher could be specified as follows (adapted from Hanushek and Rivkin 2010 and Isenberg and Hock 2010):

$$TEST_{i,g} = \lambda TEST_{i,g-1} + \alpha' X_i + \phi' S_i + \eta' T_{i,g} + \varepsilon_{i,g}$$

The test for a student i in the current grade ($TEST_g$ = eighth grade) would be a function of that student's score in the prior year ($TEST_{g-1}$ = seventh grade), a host of student characteristics captured in the X vector, a set of school and peer factors in the S vector, and the student's teacher for the year (captured in the T vector). The T column vector would have n rows, where n is the total number of teachers who taught eighth-grade math in the district. It would be zero for all rows except for one, corresponding to the teacher who taught that particular student math that year. λ , α , ϕ and η are all unknown coefficients that are estimated by fitting the model to observed test scores. The η vector contains the value-added estimates for each math teacher in the district, and the standard error for each of those estimates indicates the degree of statistical uncertainty surrounding the estimate. If the value-added coefficient for a teacher in the η vector is positive and statistically different from zero, this indicates that the teacher contributed more to his or her student's learning in a year than the average teacher in the district would have.

Prior-year test scores, which tend to be highly predictive of current-year scores, are essential to the value-added model. These scores summarize all of the student characteristics, both observable and otherwise unobservable (such as level of parental involvement in education at home), that affect a student's performance at the start of the year. The baseline test score, however, does not provide any information about the likely growth of a student's score during the year. The vector of student characteristics (X) and peer/school characteristics (S), therefore, serves the function of generating expected gains based on observable student, peer, and school characteristics.

The control variables frequently used in these models for student characteristics include whether a student is eligible for free or reduced-price lunch (indicators of poverty), the presence of specific learning disabilities, and whether a student is learning English as a second language. Some

models also control for gender or race and ethnicity, although these controls are controversial because they imply different expectations for student *gains* based on these dimensions. However, many models indicate that, after accounting for baseline and other variables, gender and/or race and ethnicity do not significantly increase explanatory power and therefore can be removed from the regressions without biasing the effectiveness estimates or making them less precise.

Common control variables for peer and school characteristics include class size and school-wide averages of the student characteristics described above, e.g., percent of a school's students who are learning English as a second language. Many models also include binary variables that indicate which school a student is in. This fixed-effects model is a flexible approach for removing the mean school effects from a teacher's value-added estimate, but it also means that the value-added estimates end up being comparisons between teachers within a school and not across schools (Lipscomb et al. 2010).

Some value-added models include multiple prior years of student test scores in the regressions. For example, a regression estimating the effectiveness of eighth-grade math teachers would use not only their students' seventh-grade math scores but also their sixth, fifth, and fourth. One approach is to add successive lagged control variables to the basic value-added model described above. Other value-added models, however, use a more involved approach of directly modeling changes in student test scores over time. Teacher effects are estimated as deflections in a student's trend in growth over time (McCaffrey et al. 2003). The underlying idea in both approaches is that, compared to using one prior-year score and available covariates, using multiple years of scores can provide a more accurate prediction of a student's gains in the current year if he or she were taught by an average teacher. The more accurate prediction then allows for a more accurate estimate of the extent to which the current teacher contributed to current gains in student scores, compared to what would have occurred under the average teacher.

Value-added models clearly do a better job of isolating teacher effects on student learning than do other methods that assess effectiveness based on the change in the percentage of a teacher's students who meet proficiency standards from one year to the next, or the change in a class's average test scores from one year to the next (Baker et al. 2010; Glazerman and Potamites 2011). Both alternative methods can penalize a teacher for having a new cohort of students who, by chance, have lower baseline scores than prior cohorts. In contrast, the value-added models are agnostic to the starting points of a teacher's students, evaluating teachers only on their ability to help students achieve gains beyond what would be predicted for each particular student.

However, many researchers argue that value-added models are still likely to produce biased estimates of teacher effectiveness (Rothstein 2010; Baker et al. 2010). One reason is that these models do not control for unobservable student characteristics (such as parental involvement) that can affect a student's growth in test scores in a year. A teacher can therefore receive a high effectiveness score simply because he or she had students with unobservable advantages that— independent from any influence of the teacher—will help them achieve higher gains than predicted. Researchers have not reached consensus on this issue, with strong theoretical and empirical arguments on both sides (Rothstein 2010; Kane and Staiger 2008).

III. APPLICABILITY OF VALUE ADDED TO ISOLATING PHYSICIAN EFFECTS ON PATIENTS' HEALTH OUTCOMES

For the value-added approach in education to be useful for measuring physician performance, the health outcomes of a physician's patients would have to meet three criteria that student test scores generally meet. First, the outcome should be measurable on a continuous scale that can be used to make meaningful distinctions in health status between most patients. Second, the outcome should be measured repeatedly (at least twice) for each patient, and the score at the start of each measurement period should be reasonably predictive of the score at the end of the period.¹ Third, a physician should be able to influence the change in the score during the measurement period.

Several types of outcome measures are likely to meet these criteria: (1) intermediate clinical outcomes measured with laboratory tests (e.g., glucose levels for patients with diabetes, cholesterol or blood pressures for patients with coronary artery disease); (2) measures of functional status or physical functioning, and (3) survey-based measures of health status, such as the SF-36 (Ware and Sherbourne 1992). In addition, composites of measures can be constructed to meet these criteria.

Hemoglobin A1C (HbA1c) levels for patients with diabetes provide an example of how an outcome measure can meet the three criteria described above:

- **A continuous measure that captures meaningful distinctions in health status.** Hemoglobin A1c measures a person's average blood sugar levels in the past two to three months. Patients with diabetes are at high risk for microvascular complications like blindness and kidney failure, and this risk increases with higher blood sugar levels. HbA1c can be measured on a continuous scale (percentage of hemoglobin in the blood that is glycated), and HbA1c scores represent meaningful distinctions in health status between patients. More specifically, the risk of blindness and other microvascular complications increases exponentially with HbA1c level; someone with a very high HbA1c level (>10 percent) is at much greater risk than someone with a high level of HbA1c (8.5 percent), who in turn is at much greater risk than someone with an optimal level of HbA1c (<7 percent) (American Diabetes Association 2012).
- **Score at one measurement period predictive of the score in the next.** HbA1c levels change gradually over time in response to changes in diet, exercise, and medical intervention, including drug therapy. These gradual changes mean that a person's HbA1c level at one point in time, such as the start of the year, is likely to be quite predictive of the value at the end of the year.
- **Physician influence on outcomes.** Finally, physicians have several tools at their disposal for lowering HbA1c levels. These include encouraging patients to adopt

¹ If a patient's initial score is predictive of the final score, it is likely that including the initial score in a value-added model will help to correct for biased estimates of a physician's influence on health outcomes. Bias occurs if two conditions are true: (1) baseline health status is correlated with final health status, and (2) baseline health status is correlated with the patients seen by a physician (e.g., some physicians tend to see sicker patients than do other physicians). If both of these conditions hold, failure to control for baseline health status will bias the value-added estimates in that they would lead one to conclude that physicians who see sicker patients have poor value-added scores.

therapeutic lifestyle changes, including HbA1c-lowering dietary and exercise options, and drug therapy, such as metformin and insulin.

For health outcomes that meet the three criteria, a value-added model similar to the one described above could be used to isolate a physician's effect on the outcomes from the effects of other variables, including illness severity, genetic predisposition to illness, and socioeconomic status (which can influence, for example, access to healthy foods). For instance, to determine a physician's influence on a patient's HbA1c levels, a value-added model could take the form of:

$$\text{HbA1c}_{i,g} = \lambda \text{HbA1c}_{i,g-1} + \alpha' \mathbf{X}_i + \varphi' \mathbf{S}_i + \eta' \mathbf{P}_{i,g} + \varepsilon_{i,g}$$

In this model, patient i 's current HbA1c level is a function of his or her HbA1c level in the prior year², plus a host of patient-level characteristics in the \mathbf{X} vector (such as age and co-morbid conditions), system-level characteristics in the \mathbf{S} vector (such as the size of the physician's medical practice or the availability of other physicians in the market area), and the specific contribution of the patient's physician for the year.³ The parameter estimates in the η vector would provide value-added estimates for each physician in the sample. A positive value-added estimate that is statistically different from zero would signal that the physician was able, on average, to lower his or her patients' HbA1c (or stem an anticipated increase in HbA1c) more than an average provider in the data set would have done.⁴

² Like value-added models that use multiple years of prior test scores for each student, this model could use multiple prior years of HbA1c scores for each patient. With these additional data, the model would generate more accurate predictions of what each patient's outcome would have been had he or she seen an average physician in a year—and therefore more accurate estimates of each physician's unique contributions to actual outcomes.

³ Determining who counts as the patient's physician for the year is complicated, given that Medicare beneficiaries often see many providers in a year (see the next section). In this model, each patient is attributed only to the doctor who provided the most evaluation and management services to that patient during the year.

⁴ Because lower HbA1c levels (down to about 6.5 to 7 percent, depending on the patient) are desirable, a good value-added score would actually be negative. In this example, it is assumed that the value-added score would be the negative of the coefficient estimate, and a positive value-added score would mean that the physician was able to lower HbA1c levels more than the average physician could.

IV. POTENTIAL BENEFITS OF USING VALUE-ADDED MODELS IN HEALTH CARE

The value-added approach to isolating physician effects on outcomes could provide three advantages for the Medicare value-based modifier.

A. Risk Adjustment

The purpose of risk adjustment is to remove the influence of patient-level risk factors on measured outcomes so that they do not bias estimates of physician performance. For the value-based modifier, CMS plans to risk adjust the health outcome measures but has not settled on a method for doing so. The agency's "jumping off point will be prior work on risk-adjusting hospital mortality and re-admission rate measures" (CMS 2010). This strategy suggests that CMS will consider adjusting outcomes on the basis of primary and secondary diagnoses recorded on inpatient, outpatient, and physician claims.

Risk adjustment based on claims data is an important starting point. However, diagnoses in claims do a poor job of capturing the *severity* of a patient's condition, and models based on these diagnoses do not control for other factors that can influence outcomes, including undiagnosed conditions and socioeconomic status. The value-added approach would control for these factors through two means: (1) the prior baseline value for the outcome (for example, the prior HbA1c level), which encompasses all the measurable and otherwise unmeasurable factors that influence the initial level of the outcome and are, in turn, predictive of the subsequent level of the outcome and (2) measured variables, such as age and comorbid conditions, that can influence the change in the outcomes over the measurement period. However, this value-added method still cannot control for unobservable characteristics that may influence the *change* in the outcome.

Appropriate risk adjustment is important to achieving physician buy-in to the program and to creating effective incentives. While over 70 percent of general internists nationwide agree in principle that physicians should be given financial incentives based on quality, an even larger fraction reported that making payments based on unadjusted measures may lead physicians to avoid high-risk patients (Casalino et al. 2007). One physician said, "If my pay depended on A1c levels, I have 10 to 15 patients whom I would have to fire. The poor, unmotivated, obese, and noncompliant would all have to find new physicians" (Casalino et al. 2007). Another influential study notes that, even after controlling for demographics and diagnostic data, providers could substantially improve their performance (as measured by the percentage of patients who met A1c thresholds) simply by dropping the one to three patients with the highest A1c readings in the prior year (Hofer et al. 1999). However, implicit in this finding are the potential benefits of a value-added approach: it would directly control for prior-year A1c levels, diminishing the incentive to drop high-risk patients.

In developing a value-added model for measuring physician performance, it is important to address the question of whether the model should adjust for the extent to which a patient adheres to a physician's treatment recommendations during the measurement period. Echoing the physician quoted above, some commenters on CMS's plans for the value-based modifier noted that physicians should not be held accountable for outcomes if patients do not adhere to prescribed treatments; a patient with advanced diabetes who does not take insulin appropriately is a good example. From this perspective, it would make sense to control for a patient's adherence by using a proxy measure like the percentage of prescribed medications that were actually filled.

However, one of the tasks of physicians is to influence patient behavior, and physicians can adopt specific practices (like scheduling follow-on appointments or thoroughly answering patients' questions) that increase the likelihood of adherence (DiMatteo et al. 1993). From this perspective, it would be inappropriate to control for a patient's adherence during the measurement period because doing so would cancel out the effects of one main pathway through which providers can influence patient outcomes. It is, however, appropriate to control for a patient's extent of adherence *before the measurement period* because a physician cannot influence past patient behaviors. A value-added model does this by definition because the prior year's measure of the health outcome (which the model includes) reflects the many factors, including prior adherence, that affect the baseline level for the outcome.

B. Rewarding Patient Improvement

The clinical outcome measures that Medicare currently plans to include in the modifier are based on whether patients achieve critical thresholds. For example, a provider would be rated based on the percentage of his or her patients with coronary artery disease who are in LDL control (<100 mg/dL) or the percentage of patients with diabetes who have A1c control (<8.0 percent). When measured at a single point in time, these measures make sense because they indicate whether a patient's treatment meets evidence-based recommendations for optimal care. However, they miss an opportunity to reward physicians who have generated substantial improvement for a patient, even if that patient's value does not cross a critical threshold. For example, CMS may want its incentive system to reward a physician who lowers a patient's HbA1c level from 10.5 to 9.5 percent, which can substantially reduce the patient's risk of adverse microvascular events like blindness even though the physician would not qualify as having met the optimal threshold for that patient. Indeed, a move from 10.5 to 9.5 percent would likely have more of a positive influence on patient health than a move from 8.5 to 7.5 percent, given the exponential relationship between HbA1c levels and adverse outcomes. A value-added approach would reward a physician for any improvement in the outcome measure that was beyond what you would expect had that patient been treated by the average physician.

C. Controlling for System-Level Characteristics

A value-added approach also gives CMS the flexibility to control for system-level characteristics that can influence patient outcomes but that are not under a physician's control, or at least not in the short run. For example, CMS may want to control for whether a physician practices in a market where there is a shortage of medical specialists, as can often happen in rural areas. These shortages, which are not under an individual physician's control, may worsen patient outcomes. However, the value-added model could control for market-level characteristics like this by including them in the \mathbf{S} vector in the model above (for example, as a binary variable for whether a physician's market area has a shortage of specialists).

CMS may also consider controlling for practice characteristics, like practice size. Large group practices are much more likely than smaller, independent offices to (1) have nurses who can teach patients self-management practices, (2) adopt evidence-based guidelines for all physicians in the practice, and (3) maintain electronic disease registries to track patients' clinical outcomes over time (Casalino et al. 2003). In the short run, an individual physician has little influence on practice size, so CMS might be able to better measure a physician's value-added by controlling for the size of his or her practice when estimating a physician's effects. In the long run, however, physicians choose the type of practice they are in, and CMS may want to encourage physicians to move into larger

practices. This objective would argue against controlling for practice-level characteristics. A value-added approach would provide CMS with some flexibility in making these decisions because the agency could choose whether to include such practice-level characteristics in the \mathbf{S} vector.

A value-added approach also appears to address one particularly difficult methodological challenge—isolating the influence of one physician from the other providers a patient sees in a year. This challenge is central to measuring value added in Medicare because the average Medicare beneficiary sees two primary care physicians and five specialists over the course of the year (Pham et al. 2007). There is a corollary to this challenge in education research: it is common for one student to be taught a particular subject by more than one teacher during the year. This may happen, for example, if the student changes schools during the year or is pulled out for individual instruction in math two times a week. Using a value-added approach, Isenberg and Hock (2010) addressed this situation through dosage models in which values in the \mathbf{T} vector are not limited to 0 and 1 (for whether a teacher did or did not teach a student that year). Rather, these values continue from 0 to 1, indicating the percentage of time (the “dosage”) that a student spent with each teacher that year (the sum of all teachers’ dosage equals 1). The resulting value-added estimates represent the individual contributions made by each teacher to a student’s gains in achievement.

The dosage model, however, would probably not work for Medicare. Specifically, it is likely that the percentage of time spent by a patient with a physician in a year (the physician “dosage”) will be an indicator of the patient’s health. For example, a person whose health deteriorates, leading to high HbA1c levels and a microvascular complication may end up seeing an endocrinologist for most of his or her care for the year—while a patient whose condition remains stable may see only a primary care doctor in that year. In this case, a dosage model would likely show that whenever the endocrinologist provided the largest dose of care, outcomes were poor. But this would probably not be because the endocrinologist provided sub-par care; rather, the poor outcomes would have driven the patient to the endocrinologist. For this reason, it is likely that value-added models would need to attribute patients to individual providers, using rules (or variations of them) that Medicare has already developed for assessing physicians’ financial performance (discussed in the next section).

V. CHALLENGES IN APPLYING VALUE-ADDED MODELS IN HEALTH CARE AND STRATEGIES FOR ADDRESSING THEM

This section describes several practical and conceptual challenges associated with using value-added models to measure physician performance and possible strategies for addressing them.

A. Measures “Top Out”

For a value-added model to work well, the outcome variable must allow us to draw meaningful distinctions in health status between most, if not all, patients in the model. However, many clinical measures cannot do this because they have a “cut-point” above which (or below which) additional increases (or decreases) do not signal better health (and may in fact signal poorer health).⁵ One strategy for addressing this challenge is to combine multiple measures into a single, global measure. For example, for diabetes, optimal care means having LDL, HbA1c, and blood pressure all within control. A continuous composite measure that brings together the scores in each of these measures is much less likely to top out than any one of the measures alone because it is harder to control all three factors separately than it is to control just one. Another strategy is to limit the value-added calculation to patients whose baseline values are out of control; e.g., patients with HbA1c levels above the normal range.

B. Long-Term Physician Patient Relationships

As mentioned, teachers tend to teach their students for one year, but physicians often see their patients for many years. A value-added model that rewards improvements in the health outcome could unintentionally penalize a physician who, through prior treatment and counseling, brought a patient’s clinical measures under control but cannot demonstrate any further improvement because there is no more room to improve. One possible solution is to divide a physician’s patients into two groups: those whose clinical measure are already under control (for a single measure or a composite) at the start of the measurement period and those whose measures are not. The value-added model could be applied just to the former group.⁶ However, even with this approach, performance measures based on a value-added model could unintentionally create incentives for a provider to shift his or her attention away from healthier patients close to being under clinical control if he or she can optimize the performance score (with the least additional effort) by focusing on sicker patients for whom similar absolute changes in clinical measures may be easier to obtain. One way to address this concern is to calculate separate value-added scores for each physician by categories of patients (e.g., those farther from versus those closer to control) and then combine the scores into a final score for that physician. Whether this additional complexity is justified depends on whether the

⁵ For example, decreases in HbA1c below 6.5 percent can, for some patients, create substantial risk of seizures due to hypoglycemia (too low blood sugar).

⁶ A physician could be rewarded for the percentage of patients who were under control at baseline and stayed in control, and penalized for patients who were under control at baseline but did not sustain it over time. A model similar to a value-added model could be applied to reward physicians who are able to keep patients under control who, based on observables like age, were predicted to be likely to slip out of control after a year. The key difference between this model and the value-added models discussed in the paper is that this model uses binary variables (in or out of control) at baseline and at the end of the measurement periods, whereas the value-added model uses a continuous measure (e.g., HbA1c level).

effort required to achieve a given increment of improvement does, in fact, depend on a patient's baseline score.

C. Attributing Patients to Providers

For a value-added model to work, individuals (be they students or patients) must be attributed to specific professionals (teachers or physicians). In the education models, students are attributed to the teachers who taught them during the year. While it can be difficult to decide on rules for attribution in some cases (e.g., how to attribute students who switch schools mid-year), one could argue that it is even more complicated in health care because patients often see multiple providers in a year. With regard to Medicare, however, CMS is already facing this challenge for other measures it would like to apply to physicians, including total patient costs. The agency has developed and tested a variety of attribution rules—e.g., attributing patients to providers who supplied the lion's share of a patient's care in a year—that could be adapted to a value-added model.⁷ An important step in testing a value-added model in health care would be to determine the extent to which conclusions about physician performance are sensitive to small changes in attribution rules.

D. Statistical Uncertainty

One of the major challenges associated with the use of value-added models in education is that there can be a wide degree of uncertainty in the estimates of a teacher's performance (Baker et al. 2010; Schochet and Chiang 2010). If unaddressed, this uncertainty could lead to major decisions that are unfair (e.g., tenure decisions based on measured performance that is substantially different from the teacher's true value-added). It could also blunt incentives by muddying the relationship between additional effort and measured performance.

It is likely that value-added estimates for individual physicians would also be imprecise, particularly if the number of patients attributed to a physician for whom the measured outcome variable is relevant is small. In education, approaches to increasing the precision of performance estimates include using test scores from multiple years of classes (in other words, increasing the sample size for the estimate); combining value-added scores with other, independent measures of teacher performance, such as principals' evaluations; and calculating scores at a higher level of aggregation (e.g., for all the teachers in a given subject or for all the teachers in a school, which, again, increases sample size). The following similar approaches could be taken in health: using multiple years of patient's outcomes; combining value-added measures with other measures of physician performance, such as their scores on clinical process measures; and calculating value-added scores for groups of physicians in a practice. The last approach would also be an incentive for physicians to achieve better health outcomes for their patients because their measured performance would depend not only on their own performance, but on that of their colleagues as well.

⁷ These attribution methods could be refined by creating separate patient attribution rules for each health outcome. Specifically, CMS could first identify the types of physicians who could reasonably be expected to influence the outcome—for example, primary care physicians and endocrinologists, but not orthopedic surgeons, could be expected to influence a patient's HbA1c levels. For each outcome, CMS could then attribute the patient to *all* physicians who could influence the outcome or to the single physician within that group who provided the most care that year.

E. Availability of Data

One reason value-added models have developed in education is that the student-level, longitudinal test data needed for the models are routinely collected in many school districts—partly because of federal testing requirements. However, Medicare maintains very limited clinical data, such as HbA1c levels, for patients. This situation is likely to change over time as physicians are required to report clinical data as part of the Physician Quality Reporting System or as a result of the meaningful use requirements for electronic health records. CMS is in a good position to collect and store these outcome data because, for every Medicare patient, it maintains detailed administrative records that can be linked over time. CMS may be able to capitalize on these growing data sets to implement value-added measures.

F. Teaching to the Test

There is a concern that test-based assessments of teacher performance in education, including value-added approaches, prompt teachers to “teach to the test,” possibly at the expense of other important topics and useful learning skills. This concern about teaching to the test grows as the subject matter covered by the test narrows, since narrow tests may encourage teachers to avoid important topics that are not covered by the test. While this phenomenon is also a concern in any assessment of physician performance based on specific quality indicators, a value-added approach could potentially exacerbate the concern by further restricting the set of quality measures that can be used to assess performance. This happens because the set of outcomes measures that are likely to meet the criteria described above is smaller than the full set of outcomes relevant to patient health. Therefore, value-added measures (and payments based on them) may distort patterns of care, diverting attention from other care that is at least, if not more, important for a patient’s overall health. Several techniques can be used to overcome this problem, including identifying additional quality measures that encompass broader health status or functional status measures, broadening performance measurement by adding process measures or other measures that do not necessarily lend themselves to value-added modeling, and creating composite measures that meet value-added criteria out of a wide array of measures that do not meet the criteria individually.

VI. CONCLUSION

Value-added models for Medicare could provide a powerful way to (1) control for patient- and system-level characteristics that influence patient outcomes and (2) reward providers for improvements in patient health that are not necessarily tied to particular thresholds such as HbA1c levels, blood pressure, and so on. However, whether a value-added approach could actually realize its potential for measuring physician performance depends on many factors. First, it would have to be vetted and accepted by physicians. Second, empirical testing would be needed to identify whether the outcomes of interest meet the criteria for a value-added approach—including the extent to which a measurement at the start of an observation period predicts outcomes at the end of the period. Third, sensitivity tests would need to show that the value-added estimates are not highly sensitive to small changes in the design of the regression models (for example, in the patient attribution rules). Finally, the conceptual and practical challenges described in this paper will need to be addressed.

While the paper has focused on individual clinical measures to illustrate the promise that value-added models may have in health care, a multidimensional measure of patient health—either obtained via surveys or by creating composite measures from electronic health records—could, in the long run, serve the same purpose as student achievement tests serve in education. More comprehensive measures of health could provide better incentives for physicians to focus on the whole person because individual measures—be they clinical process or outcome measures—may inadvertently draw a physician’s attention to aspects of quality that are less central to the ultimate goal of improved patient health.⁸

⁸ In some cases, moving from individual to more comprehensive measures of health may mean that a physician receiving the value-added score will arguably have less control over the measure. For a value-added model to work, the physician receiving the score would have to have some, although not complete, influence on the measure(s) selected.

REFERENCES

- American Diabetes Association. “Standards of Medical Care in Diabetes—2012.” *Diabetes Care*, vol. 35, Supplement 1, 2012, pp. S11–S63.
- Baker, E., P. Barton, L. Darling-Hammond, E. Haertel, H. Hodd, R. Linn, D. Ravitch, R. Rothstein, R. Shavelson, and L. Shepard. “Problems with the Use of Student Test Scores to Evaluate Teachers: EPI Briefing Paper.” Washington, DC: Economic Policy Institute, August 2010.
- Casalino, L., G. Alexander, L. Jin, and T. Konetzka. “General Internist’s Views on Pay-for-Performance and Public Reporting of Quality Scores: A National Survey.” *Health Affairs*, vol. 26, no. 2, 2007, pp. 492–499.
- Casalino, L., R. Gillies, S. Shorthell, J. Schmittiel, T. Boddenheimer, J. Robinson, T. Rundall, N. Oswald, H. Schaffler, and M. Wang. “External Incentives, Information Technology, and Organized Processes to Improve Health Care Quality for Patients with Chronic Conditions.” *Journal of the American Medical Association*, vol. 306, no. 4, 2003, pp. 434–441.
- Centers for Medicare & Medicaid Services. “Listening Session Regarding Physician Feedback Program and Implementation of the Value-Based Payment Modifier for Fee-for-Service Medicare.” 2010. Available at [\[https://www.cms.gov/PhysicianFeedbackProgram/Downloads/092410_Listening_Session_Feedback_Program_Transcript.pdf\]](https://www.cms.gov/PhysicianFeedbackProgram/Downloads/092410_Listening_Session_Feedback_Program_Transcript.pdf). Accessed October 14, 2011.
- DiMatteo, R., C. Sherbourne, R. Hays, L. Ordway, R. Kravitz, E. McGlynn, S. Kaplan, and W. Rogers. “Physicians’ Characteristics Influence Patients’ Adherence to Medical Treatment: Results from the Medical Outcomes Study.” *Health Psychology*, vol. 12, no. 2, 1993, pp. 93–102.
- Glazerman, S., S. Loeb, D. Goldhaber, D. Staiger, S. Raudenbush, and G. Whitehurst. “Evaluating Teachers: The Important Role of Value-Added.” Washington, DC: Brown Center on Education Policy, The Brookings Institution, November 2010.
- Glazerman, S. and L. Potamites. “False Performance Gains: A Critique of Successive Cohort Indicators.” Working paper. Washington, DC: Mathematica Policy Research, December 2011.
- Hanushek, E., and S. Rivkin. “Generalizations About Using Value-Added Measures of Teacher Quality.” *American Economic Review*, vol. 100, no. 2, May 2010, pp. 267–271.
- Hofer, T., R. Hayward, S. Greenfield, E. Wagner, S. Kaplan, and W. Manning. “The Unreliability of Physician ‘Report Cards’ for Assessing the Costs and Quality of Care of a Chronic Disease.” *Journal of the American Medical Association*, vol. 281, no. 22, 1999, pp. 2098–2105.
- Isenberg, E., and H. Hock. “Measuring School and Teacher Value Added for IMPACT and TEAM in DC Public Schools, Final Report.” Washington, DC: Mathematica Policy Research, August 2010.
- Kane, T., and D. Staiger. “Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation.” NBER Working Paper 14708. Cambridge, MA: NBER, 2008.

- Lipscomb, S., B. Teh, B. Gill, H. Chiang, and A. Owens. "Teacher and Principal Value Added: Research Findings and Implementation Practices, Final Report." Princeton, NJ: Mathematica Policy Research, September 2010.
- McCaffrey, D., J.R. Lockwood, D. Koretz, and L. Hamilton. "Evaluating Value-Added Models for Teacher Accountability." Santa Monica, CA: RAND Corporation, 2003.
- Office of the Federal Register. "Improvements to the Physician Feedback Program and Establishment of the Value-Based Payment Modifier." *Federal Register*, vol. 76, July 19, 2011a, pp. 42902–42913.
- Office of the Federal Register. "Establishment of the Value-Based Modifier and Improvements to the Physician Feedback Program." *Federal Register*, vol. 76, November 28, 2011b, pp. 73425–73438.
- Pham, H., D. Schrag, A. O'Malley, B. Wu, and P. Bach. "Care Patterns in Medicare and Their Implications for Pay-for-Performance." *New England Journal of Medicine*, vol. 356, March 2007, pp. 1130–1139.
- Rothstein, J. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics*, vol. 125, no. 1, January 2010, pp. 175–214.
- Schochet, P., and H. Chiang. "Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains (NCEE 2010-4004)." Washington, DC: National Center for Education Valuation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, July 2010.
- Ware, J.E., and C.D. Sherbourne. "The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual Framework and Item Selection." *Medical Care*, vol. 30, no. 6, June 1992, pp. 473–483.

About the Series

Policymakers require timely, accurate, evidence-based research as soon as it's available. Further, statistical agencies need information about statistical techniques and survey practices that yield valid and reliable data. To meet these needs, Mathematica's working paper series offers policymakers and researchers access to our most current work.

For more information, contact Greg Peterson, researcher, at gpeterson@mathematica-mpr.com, or Eric Schone, researcher, at eschone@mathematica-mpr.com.

www.mathematica-mpr.com

**Improving public well-being by conducting high-quality,
objective research and surveys**

PRINCETON, NJ - ANN ARBOR, MI - CAMBRIDGE, MA - CHICAGO, IL - OAKLAND, CA - WASHINGTON, DC

MATHEMATICA
Policy Research

Mathematica® is a registered trademark
of Mathematica Policy Research, Inc.