Linking quality and cost indicators to measure efficiency in health care (Version 3)

A paper commissioned by the National Quality Forum

Andrew M. Ryan, Ph.D.

Christopher P. Tompkins, Ph.D.

June 16, 2014

Acknowledgements

The authors would like to acknowledge the staff at the National Quality Forum, particularly Ashlie Wilbon, Taroon Amin, and Erin O'Rourke, for their contributions and assistance with this project. We would also like to acknowledge the outstanding contributions of the expert panel, particularly Joyce Dubow and Carole Flamm, the panel co-chairs. We would also like to acknowledge the research assistance of Jayme Mendelsohn.

Contents

20

Purpose of the commissioned paper

The National Quality Forum has commissioned a paper to assess alternative approaches to link – or combine – measures of quality and cost for the purpose of measuring efficiency in health care. This paper reviews various approaches—both established and novel— to measure efficiency. These include composite measures and approaches that keep the quality and cost domains separate when assessing efficiency. The paper also considers the implications of alternative methods for profiling and scoring providers based on their measured efficiency. In addition to assessing the technical issues related to measuring and profiling efficiency, we will consider the implications for using alternative approaches in the context of various programs, such as the creation of tiered insurance networks and value-based payment.

Our goal in writing the commissioned paper is to help build consensus about the key considerations and appropriateness of alternative approaches for combining quality and cost measures into quantitative measures of efficiency. This paper will serve as a foundation to inform the deliberations of a multi-stakeholder expert panel that will provide input on the methodological challenges to linking cost and quality measures and the best practices for combining cost and quality measures to assess efficiency of care.[1]

A substantial literature has also been devoted to understanding and measuring efficiency in health care.[2] While questions of efficiency in health care have been of interest for decades,[3,4] this interest has accelerated in recent years.[5] However, as identified by a recent systematic review commissioned by AHRQ, considerations of quality of care have been largely absent from this literature.[5] Instead, researchers have evaluated economic efficiency using a

variety of techniques to understand how a given output (e.g. a hospital day) can be optimized for a given set of health care inputs (e.g. physician labor, nurse labor). While the study of economic efficiency in health care is of great importance, it is not the focus of this paper.

In this paper, we seek to evaluate the specific case in which cost, borne by the payer (either the patient or the purchaser), is the input of interest and quality of care is the output of interest.[1] As such, we are interested in the assessment of efficiency only through the joint consideration of cost and quality. We will not consider approaches to the measurement of efficiency– such as brand prescribing rates or rates of MRI for patients with back pain – that seek to identify relative resource use and appropriateness.[6] Measuring inappropriate resource use, or "waste", clearly has value but represents an overly narrow interpretation of efficiency.[6] The use of health care services that are *never* clinically indicated are very limited and account for a small amount of health care spending.[7] For this reason, focusing simply on reducing "wasteful" resource use is unlikely to substantially reduce health care spending, or increase efficiency. There is also a large literature concerned with the relationship between costs and quality,[8-11] and a smaller literature on relationship between economic efficiency and quality.[12] While relevant to the concept of efficiency that we seek to understand, this literature is not primarily concerned with profiling individual providers on the basis of efficiency.

[1] While non-financial costs borne by patients (e.g. opportunity costs and travel costs associated with treatment), and administrative costs borne by providers when interacting with payers are important, consideration of these costs is beyond the scope of current efforts to profile the efficiency of health care providers.

Key Definitions

This project will reference a number of common terms that may have different connotations for different audiences. Throughout this project, we will apply a modified version of the definitions from the National Quality Forum's Patient-Focused Episodes of Care project:[13]

Quality of care: the degree to which health services for individuals and populations increase the likelihood of desired health and patient experience outcomes and are consistent with professional knowledge[14]

Cost of care: measures total health care spending, including total resource use and unit price(s), by payor or consumer, for a health care service or group of health care services associated with a specified patient population, time period, and unit(s) of clinical accountability.

Efficiency of care: measures the cost of care associated with a specified level of quality of care. "Efficiency of care" is a measure of the relationship of the cost of care associated with a specific level of performance measured with respect to the other six IOM aims of quality.

Value of care: measures a specified stakeholder's (such as an individual patient's, consumer organization's, payor's, provider's, government's, or society's) preference-weighted assessment of a particular combination of quality and cost of care performance.

As used in this project, the terms efficiency and value correspond to the respective definitions adopted previously by NQF and other stakeholders. Using these definitions, efficiency can be

assessed objectively. By profiling providers' quality, cost, and efficiency, and showing the component pieces, it is reasonable to assume that efficiency can be measured and displayed in a way that allows stakeholders to consider "value" as a preference-weighted assessment of the component pieces; i.e., quality, cost, and efficiency. For example, one approach might determine a provider to be "high quality," while also "high cost," based on its performance in relation to averages in both dimensions. An alternate approach is to insert an intermediate step of measuring efficiency. This might conclude that the provider is "high quality," but actually "low cost" when measured only against providers with similarly high quality, and therefore has high efficiency. Stakeholders can make value inferences in either case. The intermediate step serves to clarify the process by making explicit the objective relationships between quality and cost from which general and specific subjectively-weighted inferences are made regarding value.

93

Section 1. Why combining quality and cost measures to measure efficiency in health care matters

Improving the efficiency of health care delivery in the United States is critical. Recent attempts at payment reform, such as pay-for-performance and public quality reporting, have failed to reduce cost growth.[15,16] By focusing primarily on quality measures of underuse – such as non-adherence with evidence-based care – these programs have not provided direct incentives for increased efficiency. Previous efforts to reign in cost growth through managed care, such as capitated payment and utilization review, focused primarily on reducing costs rather than improving quality of care.[17]

To address these shortcomings, the Patient Protection and Affordable Care Act created numerous initiatives that are intended to improve the *efficiency* of US health care –not quality or cost alone. These initiatives include the Physician Value-Based Payment Modifier,[18] Hospital Value-Based Purchasing,[19] The Medicare Advantage Quality Bonus Program,[20] Accountable Care Organization programs,[21] and the End-Stage Renal Disease pay-for-performance program. More directly, legislation was introduced in 2009 to replace the standard update to physician payments with a geographically based "value index," which would adjust payments to physicians according to their relative quality and cost.[22]

On the private side, a number of insurers have developed products with tiered networks that are based on measures of efficiency. These products are structured to increase patient cost-sharing for using providers that are designated in a lower-efficiency tier. The first generation of these programs established tiers based almost exclusively on costs.[23] However,

insurers have developed a range of increasingly sophisticated approaches to combine indicators of cost and quality to categorize the efficiency of providers. These efforts are related to the rise of high-deductible health plans and consumerism. Patients need both quality and cost information in order to make informed choices about the services they need and the providers they should use. In addition, given the price sensitivity to plans currently sold in insurance exchanges created through the ACA,[24] insures may adopt narrower networks in order to compete on price.[25] This will likely increase insurers' use of tiered networks based on measures of provider value. Other promising private sector efforts, such as reference pricing,[26] will likely need to explicitly integrate provider quality measurement to gain greater acceptance.

These reforms require both quality and cost performance to be measured and assessed together. These ongoing initiatives share a common set of goals: 1) To better identify high and low efficiency providers and 2) To foster incentives for providers to improve efficiency. Broader efforts to better identify the relative value of health care services are related, but rely on a different set of tools and policy measures. While cost effectiveness and comparative effectiveness research seeks to understand the relative cost and effectiveness of medical treatments, efficiency profiling seeks to understand the relative efficiency of health care *providers*.

However, the desire to use efficiency measures has outpaced scientific consensus about how best to incorporate these measures into accountability efforts. As shown in section 2 of this paper, this lack of consensus for combining cost and quality measures can be seen by the disparate use of measures of efficiency across the public programs. Also, while many of the

private payer efforts to combine quality and cost have similar features, they differ in important ways.

Efforts are moving ahead to measure and profile health care providers' efficiency without a clear sense of the best approach to do so. The issues surrounding combining quality and cost measures are certainly challenging: one recent report described the state of efficiency measurement as "woefully inadequate."[27] Two high profile efforts tasked with grappling with these issues failed to recommend a strategy to do so.[28] Now is the time to develop a framework to identify the trade-offs between alternative approaches to combine quality and cost indicators in order to guide the future development, evaluation, and use of efficiency measurement in health care.

Section 2. Options for combining quality and cost measures

Methods for environmental scan

We performed an environmental scan to identify existing approaches that were currently in use by Medicare, private payers, and other program sponsors that combine indicators of quality and cost measures to assess efficiency. We also identified novel approaches that link quality and cost indicators that are not currently in use by a program sponsor but have been developed by researchers. To be included, an approach must assess cost as an input and one or more measures of quality as the output.

We searched the PubMed databases for published articles in the English language that appeared in journals between January 1990 and April 2014. Search terms included "quality", "measuring," and "cost." We searched the bibliographies of retrieved articles looking for additional relevant publications. We then searched Google Scholar, the Cochrane Database, and conducted other general internet searches for the same search terms. This provided resources that were not limited to peer-reviewed journals. We also identified applications outside of health care that combine indicators of quality and cost (e.g. Consumer Reports "Best and Worst Cars for the Money" and US News and World Reports "Best Value Schools"). See Appendix A for information on these efforts.

Additionally, we solicited information from the National Quality Forum's Expert Panel on Linking Cost and Quality. The materials referred to us by the expert panel frequently led to the discovery of additional approaches. From the panel, we also obtained detailed information on approaches that we knew had been initiated (for instance, in Medicare).

After identifying all of the programs that simultaneously assessed quality and cost, as well as approaches proposed by researchers, we identified and described a set of mutually exclusive approaches that combine quality and cost measures to measure efficiency. We then described the basic features of these approaches. Next, we identified the programs that have used quality and cost indicators to profile the efficiency of providers. This includes programs that are currently running as well as those that are now defunct. For these programs, we obtained information on several parameters: the name of the program, the services evaluated (e.g. hospital only, physician only, all services), the level of attribution (e.g. hospital, physician practice, individual physician), the specification of quality, the specification of cost, and the approach used to combine quality and cost indicators.

Approaches used to combine quality and cost measures

We identified seven approaches that are currently in use or have been proposed by researchers to combine quality and cost indicators to measure efficiency.

The conditional model: This approach, described by Timbie and Normand as the "Univariate" approach[29] and by Tompkins et al. as the "Net-Incentive Payment Model"[30] assesses efficiency as the conditional combination of quality and cost. The approach proceeds in four steps: first quality is assessed either by a single indicator or by a composite measure; second cost is assessed, typically by a single measure of total costs; third, either or both of the quality and cost domains are classified into performance groups – frequently as "low", "average", or "high" – using specified criteria; fourth, the quality and cost classifications are combined to assess efficiency. A common approach is to define high efficiency providers as those that are classified as both high quality and low cost.

Alternatively, the Net-Incentive Payment Model assesses the difference in costs between providers within the same quality grouping. The Conditional Model is widely used by private payers to create tiers of providers based on their efficiency.

The Quality Hurdle Model and Cost Hurdle Model: A variation on the Conditional Model is the Quality Hurdle Model. This model follows the first three steps of the Conditional Model. Then, providers are subject to a minimum quality standard, the hurdle, before their cost performance is assessed. After meeting this minimum quality standard, providers may be judged on cost performance alone or may be evaluated based on their combination of quality and cost performance. A variation on the Quality Hurdle Model is the Cost Hurdle Model. Here, providers are evaluated on quality performance only after meeting a cost standard, which is typically defined as having costs that are below a specified growth rate. Hurdle Models are commonly used for shared savings programs.

The Unconditional Model. The unconditional model follows the first two steps of the Conditional Model. Then, the quality and cost domains are assigned weights and combined into a single metric. Thus, in the Unconditional Model, quality and cost are scored independently and then combined. This is the model currently used by Hospital Value-Based Purchasing.

The Regression Model: The regression model, proposed by Timbie and Normand,[29] profiles provider quality while conditioning on cost. While it is conceptually similar to the Conditional Model, it has the advantage of using regression analysis to account for the within-provider correlation between quality and cost outcomes. In contrast, the approach

taken by the Conditional Model does not account for any correlation between the quality and cost domains. The regression model is not currently used by any program sponsor.

The cost-effectiveness model: The cost-effectiveness model, proposed by Timbie and Normand,[29] differs from the other approaches in that it assigns a dollar value to the patient benefits accrued from quality domain. By doing so, this approach can dramatically change efficiency profiles. For instance, using the Unconditional or Conditional Model, a hospital with excellent mortality outcomes may be classified as having only moderate efficiency if it also has high costs. However, if the benefit of increased survival is appropriately valued and the absolute cost differences between this hospital and others are not great, this high cost hospital may in fact have excellent efficiency: it is producing desirable health outputs at a lower cost than other hospitals. A similar approach towards efficiency measurement was developed by Kessler and McClellan to evaluate the cost-effectiveness not of individual providers, but of the characteristics of hospitals.[31]

The Data Envelopment Analysis or Stochastic Frontier Analysis Model: This approach is used to identify the efficient production of quality across all observed levels of cost.[32,33] The efficient frontier is modeled and providers' efficiency can then be evaluated based on their distance from the efficient frontier. One of the key advantages of this approach is that it allows efficiency to be evaluated across continuous measures of cost and quality. It therefore does not require classification of providers into categories based on what may be arbitrary threshold values, a shortcoming of other approaches. This approach has been widely used in academic research to assess economic efficiency in health care, although almost exclusively in cases in which the output of interest is something other than quality

of care.[12] This approach is not currently used by any program sponsors to evaluate provider efficiency.

The Side-by-Side Model: This approach does not combine the quality and cost domains in any way. It follows the first two steps of the Conditional Model, then concludes by displaying the results in summary form. This model typically emphasizes the clear and intuitive display of indicators of quality and cost (e.g. star ratings). However, by leaving the specific combination of cost and quality unspecified when assessing efficiency, this model leads directly to value estimations by stakeholders.

Programs using cost and quality measures to assess efficiency

Exhibit 1 describes identified programs that link indicators of cost and quality to measure efficiency. We describe the characteristics of 24 programs for which we were able to obtain sufficiently complete information.

Of these programs, 11 profiled physicians or physician practices, 4 profiled hospitals, 3 profiled both physicians and hospitals, and 6 profiled health systems or health plans. To combine quality and cost indicators, 4 of the identified approaches used the Conditional Model, 5 used the Unconditional Model, 5 used the Side-by-Side Model, and 8 used the Quality Hurdle or Cost Hurdle Model.[2] The method used to combine quality and cost indicators was unclear for 2 programs.

[2] While Veterans Affairs hospitals use stochastic frontier analysis to profile the efficiency of hospitals, assessment of efficiency does not consider quality of care as an output.

Section 3. Illustration of models to combine indicators of cost and quality

We illustrated the implementation of several of the models to combine quality and cost measures to provide a clearer idea about their similarities and differences. To do this, we downloaded data on hospital cost and quality from the May 2, 2014 release of Hospital Compare (www.medicare.gov/hospitalcompare). Our measure of cost is Medicare Spending per Beneficiary (MSPB), an NQF endorsed measure (NQF #2158). The measure captures price-adjusted Medicare spending for all services (inpatient, outpatient, home health, hospice, skilled nursing, and durable medical equipment) for acute care hospitals for all admissions in the 3 days prior to admission and 30 days after discharge. We specified cost using the ratio of the national total spending per episode to individual hospitals' total hospital spending per episode. A higher value indicates higher cost performance (i.e., lower cost relative to the national average).

The measure of quality is the Total Performance Score from Hospital Value-Based Purchasing. The Total Performance Score is a composite measure capturing hospital quality performance related to clinical process performance (45%), patient experience (30%), and outcome performance (25%). The measure incorporates both quality attainment and quality improvement. Higher scores indicate higher quality performance.

We merged cost data from 3,260 acute care hospitals with quality data from 2,728 hospitals. Our analytic sample was 2,728 hospitals. Before combining indicators, we

standardized the quality and cost indicators by subtracting the mean and dividing by the standard deviation. The distribution of the quality and cost measures are shown in Exhibit 2.

We linked quality and cost measures to measure efficiency using the following models:

1. **The unconditional model**: The unconditional model linked quality and cost measures through a weighted combination of measure scores. We calculated two separate versions of the unconditional model, one using 70% quality and 30% cost, the other using 30% quality and 70% cost.

2. **The conditional model**: The conditional model linked quality and cost by assessing cost performance for a given level of quality. We calculated two separate versions of the conditional model that varied with respect to the precision of the quality groupings. The first version classified hospitals into terciles of quality performance and then classified hospitals into cost tritiles: low (bottom 25%), average (middle 50%), and high (top 25%) cost performance. In the second version, hospitals were classified into quality tritiles, and then classified into cost tritiles within each quality tritile. In the second model, we assigned an efficiency score of "9" (the highest score) for the top quality and top cost tritile, decreasing to "1" for the bottom quality and bottom cost tritile.

3. **The quality hurdle model**: The quality hurdle model linked quality and cost measures by setting the quality hurdle at the 25th percentile. Below the 25th percentile of quality, hospitals received an efficiency score of 0. Above the 25th

percentile of quality, hospitals' efficiency was determined based on their cost performance.

4. **The cost hurdle model**: The cost hurdle was similarly set at the 25th percentile. Below the 25th percentile of cost performance, hospitals received an efficiency score of 0. Above the 25th percentile of cost performance, hospitals' efficiency was determined based on their quality.

5. **The stochastic frontier model**: The stochastic frontier model linked quality and cost measures by estimating quality as a function of cost. Efficiency was then assessed based on hospitals' "technical efficiency", a measure of hospitals' distance from the frontier.

We did not illustrate the linking of cost and quality using the side-by-side model, because this model does not formally combine measures of cost and quality. We also did not link cost and quality measures using the regression model or the cost-effectiveness models because these models require patient-level data.

Exhibit 3 shows the hurdle models, **Exhibit 4** shows the unconditional models, **Exhibit 5** shows the conditional models, and **Exhibit 6** shows the stochastic frontier model. For each of these models, greater cost performance denotes lower cost. Hospitals toward the top right of the scatter plot have higher efficiency while those toward the bottom left have lower efficiency. The precise determination of efficiency depends on the model used to combine quality and cost indicators.

Exhibit 7 shows a correlation matrix between the efficiency scores generated from the alternative models. It indicates a low to high degree of correlation between the efficiency scores generated from the different models. The quality hurdle model has a relatively weak correlation with the other models (with the exception of the unconditional (30% quality) model r=0.78). The cost hurdle model is most strongly correlated with the unconditional (70% quality) model (r=0.81), the conditional (r=0.76), and the frontier model (r=0.87). The unconditional (70% quality) model is also highly correlated with the unconditional (30% quality) (r=0.66) and the frontier model (r=0.95), while the conditional model is strongly correlated with the frontier model (r=0.88). Together, this analysis indicates that the alternative approaches generate meaningfully different efficiency signals. This has important implications for efficiency profiling using these models.

This analysis also gives a sense of some of the pros and cons of different methods for profiling. For instance, so long that quality performance does not re-enter efficiency profiles after the hurdle is exceeded, the quality hurdle model places much greater emphasis on costs, rather than quality, when assessing efficiency. This can be seen by its correlation with the unconditional model (30% quality). The opposite is true for the cost hurdle model. The analysis also highlights that, while the creation of efficiency tiers is straightforward with the conditional model, deriving nominal efficiency scores from the conditional model requires a separate scoring system that assigns a value to conditional cost and quality performance. Tompkins and colleagues[30] propose one method to do this, but others are possible.

In the analysis of cost and quality data available on Hospital Compare, lower cost is associated with lower quality: a 1% increase in cost performance (lower costs) is associated

with a 0.19% decrease in quality performance ($p<.01$). Nonetheless, the analysis indicates that it is possible for hospitals to have both excellent quality performance and excellent cost performance: there are a number of hospitals that are close to two standard deviations higher than the mean for both quality and cost performance. In other circumstances, there may be greater trade-offs between improving quality and increasing costs. In such cases, program sponsors should accommodate their expectations to the reality of cost and quality trade-offs.

To further illustrate this point, **Exhibit 8** shows the cost per beneficiary and quality scores from a hypothetical sample of hospitals. The vertical axis is spending per beneficiary and the horizontal axis is the hospital's total quality score. The quality scores are expressed here from 0 to 1, with 0 being the lowest quality and 1.00 being the highest. Contrary to the specification of costs to illustrate the alternative models to combine quality and cost, in this example, higher levels of cost indicate worse cost performance. A trend line has been fitted to the data.

As can be seen from **Exhibit 8**, there is a slight positive correlation between cost and quality for these hospitals. This is not to say that cost and quality are slightly positively correlated for all hospitals presently or that this relationship will continue in the future. As the health care system evolves and our ability to measure quality improves, cost and quality may very well become negatively correlated. Moreover, the nature of the relationship between resource requirements and quality may vary across dimensions of quality. For example, improving certain outcomes or adhering to best practices may result in greater resource requirements, suggesting the positive correlation. Meanwhile, quality improvements in patient safety and medical errors may result in lower costs from complications and treatment failures,

resulting in a negative correlation between specified levels of quality (patient safety) and total cost of care (including complications and additional services) (Exhibit 9). Similarly, more extensive substitution of hospice and palliative care for higher-cost, marginally futile treatment approaches may have corresponding improvements in patient experience.

After calculating objective efficiency based on principles and empirical calculations, a user could then determine what value to place on that efficiency score based on subjective-preference weighting.

Exhibit 10 provides an illustrative example of how to value hospital performance under a star rating system. The Total Quality Score (horizontal axis) and the efficiency score (vertical axis) are used to assign the value scores (i.e., determine the number of stars). Note that the same efficiency score is valued differently depending on the total quality score: higher total quality results in a greater value (number of stars) for the same efficiency score. Such a star rating system might be suitable for public reporting.

Once the assessment (i.e. number of stars) of the hospital performance has been made, it could be quantified by adjusting a hospital's Total Quality Score (0 to 100 points) upwards or downwards depending on its efficiency rating. An illustrative example is provided in **Exhibit 11**.

Section 4. Summary of findings from environmental scan

Our environmental scan and illustration of alternative models for combining quality and cost indicators highlights a number of key issues related to measuring efficiency in health care.

First, there are numerous extant approaches and no clear consensus about best practices. Of the 24 identified programs, we documented five broad approaches to combine quality and cost indicators. There is considerable variation within these approaches. Many of the quality measures included in the quality domains are exclusively measures that are endorsed by the National Quality Forum or by professional societies. The cost measures used to assess efficiency, however, have generally not been endorsed by the National Quality Forum.

Interestingly, the measure sets used to assess quality for many of the approaches taken by the private payers are more expansive than those used by the public payers. For instance, many of the private efficiency efforts profile specialist physicians, who have been largely ignored by public programs. The purpose of efficiency measurement is also different in the public and private efforts: the public efforts seek to use efficiency measurement to adjust provider payments whereas the private efforts use efficiency measurement to create tiered networks or for shared-savings programs.

The alternative approaches used to combine cost and quality measures have a number of pros and cons. The Conditional Model, the Unconditional Model, the Side-by-Side Model, and to a lesser extent the Hurdle Models all have the benefit of being relatively easy to understand. (Many of the program sponsors emphasized the importance of transparency, describing efficiency measurement in simple terms on their website but also publishing detailed

methodology reports.) However, these approaches suffer from two separate aggregation problems that may undermine their validity. First, quality is almost always defined using multiple measures, and some kind of weighting scheme is required to summarize the performance of providers on these measures. The opportunity model, in which weights are based on the number of patients that are eligible to receive a given measure, remains a common approach to creating composite measures of quality. Another approach, used by the Alternative Quality Contract, assigns triple the weight to outcome measures relative to process measures. Both of these approaches to weighting measures, however, are largely arbitrary. A recent paper found that among 13 commonly used quality indicators, 7 of them accounted for 93% of the benefits to population health.[34] If weights assigned to individual performance measures do not reflect their importance to the health of patients, weighting schemes will, at a minimum, obscure the signal between observed quality and patient health.[35]

Second, as previously described, efficiency measurement has the potential to reach erroneous conclusions about the relative efficiency of providers when the relationship between measured quality and patient health is not well defined. If quality is measured by patient survival, then small improvements have the potential to yield large efficiency gains, even at large costs. However, if quality is measured by a series of measures that have little relationship with improved patient health, large improvements may not yield efficiency gains, even at small costs. [29]

Among existing programs, there is a divergence in the practice of price standardization. The public programs (Hospital Value-Based Purchasing, the Physician Value-Based Payment

Modifier, and the ACO programs) standardize payments when measuring efficiency. The private plans vary with respect to price standardization, but tend not to standardize prices.

Variation in the prices of health care services charged by different health care providers, particularly among private payers, is well documented.[36] Variation in prices among private payers is driven largely by the result of negotiations between private payers and individual providers. Measures of health care spending (i.e., cost) that do not first standardize prices will measure costs as the product of price and the quantity of services for individual providers. Measures of spending that standardize prices substitute individual provider prices with average prices across the population of providers. The decision of a program sponsor to use either unstandardized or standardized prices depends on the use case, particularly the needs of the end user. Individual patients are likely to care more about out-of-pocket spending, which will be more closely related to unstandardized prices. Private insurers that are using tiered benefits designs to encourage patients to get care from lower-priced, higher quality providers are also likely to prefer unstandardized prices as well. This highlights the importance of not "stripping out" variation that is meaningful for consumers and program sponsors through price standardization. On the other hand, program sponsors with well-established reasons for price variations (e.g. Medicare's indirect medical education and disproportionate share payments to hospitals) may wish to highlight differences in resource use, and therefore use standardized prices.

There also appears to be a general ambivalence on the part of program sponsors with respect to harmonization within the quality and cost domains. This includes harmonization of the quality and cost domains for the same populations of patients (i.e., cost is often assessed

for all patients while the quality measures apply to a narrower set of patients), for the same time intervals of measurement (i.e., the quality measures were assessed over much longer time windows than the cost measures), and the methods used to risk adjust for cost and quality outcomes (e.g. Hospital Value-Based Purchasing uses different approaches for quality and cost).

Over time, efficiency profiling appears to have shifted away from hospitals and towards profiling the efficiency physicians and physician practices. The early efforts in efficiency profiling focused on hospitals,[23] but many now profile physicians and physician groups. This may have to do with the increase in ambulatory measures and advances in physician attribution methodology but may also reflect the increased bargaining power of hospitals.

Importantly, for the examined approaches for combining quality and cost measurement, there is virtually no assessment of the reliability and validity of efficiency measurement.[5] In almost all cases, a single measure of efficiency is not defined. Instead, efficiency is defined through the joint consideration of quality and cost, with classification typically based on threshold values for both scales. While there is widespread recognition of the small sample size problem associated with efficiency measurement, the most common solution is to use a sample size cut-off as an exclusion criterion for providers' data to be profiled. Outside of Hospital Value-Based Purchasing, Bayesian reliability adjustment is not used to increase the reliability of efficiency measurement, although Leapfrog has used reliability adjustment for some surgical mortality measures. [37]

Section 5. Combining indicators of quality and cost for different use cases

Indicators of quality and cost could be combined for a variety of "use cases." Potential use cases include public quality reporting, pay-for-performance, network design, and internal efficiency profiling and improvement. The key question is whether and how the criteria for selecting models to combine quality and cost indicators may depend on a specific use case. What are the trade-offs that one might consider in selecting a model for a specific purpose? The following are some principals that could be applied to combining quality and cost indicators for different use cases:

1. When measuring efficiency, neither the cost nor quality signals should be obscured. Therefore, provider-level profiles of efficiency should show indicators of cost, quality, and efficiency side-by-side. This is particularly relevant for public quality reporting but is also recommended to ensure transparency for pay-for-performance, network design, and internal profiling and improvement. Consumers and referring providers typically make highly subjective and idiosyncratic choices about which treatments consumers should receive from which providers. Displaying measures of cost and quality can provide stakeholders with inputs to their own subjective and implicit preference-weighted decisions case-by-case. When making treatment decisions, consumers can supplement these objective measures with past experience, familiarity, convenience, and informal advice from trusted sources. In situations involving terms of regulation or contracting, the disclosure of individual measures in all relevant domains allows

stakeholders to understand the individual components which also should be disclosed for transparency.

2. The choice of the model to combine measures of quality and cost should depend on the aims of the use case. Efficiency scores and profiles should be developed and displayed across the entire relevant range of specific levels of quality. Quality and efficiency can be measured continuously or discretely. If discrete measures are used (e.g. a star system), categories should reflect meaningful differences across providers rather than arbitrary classifications based on distributinos (e.g. centiles).

3. Models that combine indicators of quality and cost differ with respect to the relative weight or importance that they place on quality and cost. For instance, the quality hurdle model places greater emphasis on cost performance, while the cost hurdle places greater emphasis on quality performance (see **Exhibit 3** and **Exhibit 7**). The choice of model used to combine quality and cost measures could have a significant impact on the relative importance of incentives to reduce costs or improve quality. Generally, failure to distinguish differences in performance in all cases above or below a hurdle or threshold correspondingly reduces incentives for achieving better performance within such wide ranges. Therefore, policymakers and stakeholders should carefully consider how the choice of model to combine quality and cost measures best meets the goals of the use case.

4. Whenever possible, continuous measures of efficiency are preferable to arbitrary classifications, particularly classification based on rankings. For some applications, such as network design, discrete classifications are necessary in order to group providers into

different network tiers. However, discrete classifications add to measurement error by grouping heterogeneous providers in homogenous groups. To avoid the potential issue of false precision introduced by the use of continuous scores, variance estimates (such as confidence intervals) should be used whenever possible. Classifications based on rankings (e.g. percentiles) have the potential to magnify the importance of small differences in efficiency if scores are clustered close to threshold values. In some cases, this problem can be addressed through measure selection, i.e., by excluding quality measures that are "topped out;" (i.e., average scores close to the theoretical maximum performance level).

5. When efficiency measures are incorporated as part of public reporting programs, program sponsors should adhere to best practices for the display of information.[38] Likewise, program sponsors should incorporate efficiency measures into pay-for-performance programs using best practices for program design.[20]

Section 6. Implications for the National Quality Forum measure endorsement process

To date, few stand-alone measures are being used to assess efficiency. Instead of endorsing specific efficiency measures, the National Quality Forum could choose to endorse a process to combine quality and cost indicators to measure efficiency. A number of measure-developers have already developed detailed processes to measure efficiency that they could submit for National Quality Forum endorsement.

If the National Quality Forum decided to endorse approaches to efficiency measurement it could consider a number of guidelines. First, the National Quality Forum could stipulate that the quality and cost measures used to measure efficiency should have been previously endorsed. If not, the developer would have to provide a compelling reason. Second, the National Quality Forum could provide guidance with respect to whether specifications of quality and cost measures should be harmonized. This may result in the modification of the specifications of measures that have previously been endorsed by the National Quality Forum. Third, the output of the efficiency measures should meet the standards of scientific acceptability established by the National Quality Forum. Specifically, efficiency classifications should be reliable and valid, and statistical testing should be able to demonstrate this.

When developing efficiency profiles of health care providers, the NQF also could recommend a specific process, or set of stages, that program sponsors could follow. The stages could include: 1) articulating the use case; 2) selecting cost and quality measures; 3) determining whether and what type of composite measures will be used to measure the cost

and quality domains,; 4) combining the quality and cost indicators in a manner that is most appropriate for the given use case.

1. **Identifying the use case**. Issues related to combining cost and quality measures for different use cases were described in the previous section. In some instances, the same measures and models to combine indicators may be used for multiple use cases (e.g., pay-for-performance and public quality reporting). In these examples, it is important make sure that the choices made for measuring efficiency are appropriate for each of the multiple use cases.

2. **Selecting cost and quality measures**. The individual cost and quality measures that are combined to measure efficiency should be reliable, valid, useable, and feasible to collect. The type of quality measure (structure, process, outcome, or patient experience) used should depend on the use case. There is a place for process measures, which often are useable and actionable, but they should be proximal to an outcome. For cost measures, it is important to anticipate the perspective of the decision-maker. Third-party payers are concerned with payments for covered services related to the particular focus of measurement, which may include broad classes of care such as ambulatory surgeries, inpatient admissions, or primary care management of various acute and chronic illnesses. For example, a health plan would evaluate the efficiency or value of surgical procedures based on formulaic or negotiated payment rates for facility and professional services (separately or bundled). A consumer perspective would focus on out-of-pocket payments for deductibles, coinsurance, and copayments for the full episode of care. Generally, providers are not identical or necessarily similar in their

relative quality, cost, efficiency, or value across lines of service; hence, the NQF framework for measuring resource use differentially for specific patient-focused episodes of care.

Whenever possible, it is preferable to harmonize the specifications of the cost and quality indicators used to measure efficiency. This includes measuring cost and quality for comparable populations of patients, for the same time intervals of measurement, and the methods used to risk adjust for cost and quality outcomes. Optimally, this would be done around common episodes. However, it often may not be possible or reasonable to harmonize cost and quality measures given prevalent limitations in current measures. One key reason for that is composite measures are often used to measure quality, and the individual measures contained in composite measures often have different data capture periods, and apply to different populations. Nonetheless, this is a principle to strive for in future measure development.

When combining measures of quality and cost to assess provider efficiency, it is essential that risk-adjustment procedures are appropriately implemented to hold variation in patient severity constant across providers. Standard risk-adjustment procedures use "indirect standardization" in which regression analysis using the entire sample of patients is used to assign severity weights to individual comorbidities and risk factors. These weights are then used to calculate the ratio of "observed" (or "predicted" [39]) outcomes, as well as "expected" outcomes,[40] and combine these to evaluate providers' outcome performance while holding risk constant. However, this approach may be problematic if samples do not overlap on risk factors. In this case, provider

outcome profiles may be confounded by specific characteristics that are unrepresented in other providers. Under these circumstances, "direct standardization" may more effectively account for differences in severity across providers.[41]

3. **Determining whether composite measures will be used to measure the cost and quality domains, and which composite approach will be used**. Some committee members voiced concern that too much information or detail on the cost and quality signals separately is lost in the composite. However, while composites may be overly complex and impractical they may also have important uses. For instance, consumers may prefer a single score that is easy to interpret, and program sponsors may need a single score to evaluate providers (e.g. for pay-for-performance). There are a number of approaches to create composite measures. These include all-or-none composites (requiring that a patient receive all recommended care for the composite to be met), composites based on opportunities (equal to the sum of successfully achieved processes of care divided by the opportunities to provide recommended care), and composites that assign different weights to different types of measures (e.g., weighting outcome measures more heavily).[42] There are also a number of NQF-endorsed composite measures. While it is preferable for programs to use NQF-endorsed composite measures, the dearth of these measures make this unreasonable in most cases. Instead, programs should have a reasonable justification for the weighting of individual measures, including the known correlation between measures and patient outcomes.[34]

4. **Combining the quality and cost indicators using the most appropriate model for the given use case**. Side-by-side displays of measures, for example for public reporting, can

include measures that are scored on mastery, rather than relative performance or rankings. For example, if a large majority of providers have similar or even identical scores on a measure, it may be informative and reassuring for consumers to be aware that their options are similar, or possibly all excellent, on that measure. Distinguishing relative efficiency or value, on the other hand, requires "grading on the curve," with meaningful underlying differences that are measured reliably.

List of Exhibits

625

626 **Exhibit 1.** Summary of programs that combine quality and cost indicators to measure efficiency[3]

	Name of program	Services evaluated	Level of attribution	Specification of quality	Specification of cost	Approach to combining quality and cost
1.	Aetna Aexcel [45]	12 categories of specialty services[4]	Specialist and physician practice level	Volume (at least 20 episodes in the last year) clinical performance structure measures (use of technology, certification) completion of performance-based improvement module claims based measures (HEDIS, readmissions, in-hospital complications)	All costs attributed to specialists for specific episodes of care	Variation on Quality Hurdle model. Quality and volume are assessed first. If costs are lower than threshold based on peer performance, providers are designated for Aexcel network.
2.	Anthem Blue Precision	5 categories of specialty services[5]	Specialist and physician practice level	Receipt of designation from National Committee for Quality Assurance (NCQA) or Bridges to Excellence, or performance on clinical process measures.	Combination of all attributed costs, diagnostic testing, prescription use, procedures and follow-up care, and hospital care.	Quality hurdle model. Physicians must first be recognized for quality. Physicians are then designation for recognition if their cost performance is not significantly higher ($p <$ .10) of 110% the geographic average.
3.	Blue Cross and Blue Shield Blue	6 categories of specialty services[6]	Hospital	Nationally consistent measures based on structure,	All costs for specific episodes of care	Quality Hurdle Model

[3] In addition to the programs identified in this table, we are aware of a number of other programs that appeared to combine indicators of quality and cost to measure efficiency. These include programs initiated by Castlight Health, the Minnesota Smart Buy Alliance, PacifiCare, the Puget Sound Health Alliance, Blue Shield of Oregon, Tufts Health plan, and the Wisconsin Department of Employee Trust Funds. However, we were unable to obtain detailed information about how the programs were specified.

[4] Cardiology, Cardiothoracic surgery, Gastroenterology, General surgery, Neurology, Neurosurgery, Obstetrics and gynecology, Orthopedics, Otolaryngology/ENT, Plastic surgery, Urology, and Vascular surgery

[5] Rheumatology, cardiology, obstetrics/gynecology, endocrinology, and pulmonary medicine.

[6] Six specialty care areas are included: Spine Surgery, Knee and Hip Replacement, Cardiac Care, Transplants, Bariatric Surgery and Complex and Rare Cancers. The three specialty care areas with asterisks have Blue Distinction Centers; Blue Distinction Center+ designations will continue to roll out in other areas, beginning with Transplants in early 2014.

	Name of program	Services evaluated	Level of attribution	Specification of quality	Specification of cost	Approach to combining quality and cost
	Distinction Centers® for Specialty Care[46-48]			process, outcomes, and patient experience. Hospitals must meet quality thresholds for each domain. Measures were developed with input from the medical community.	(including facility, professional, other). Each provider's cost of care is calculated on an episode basis, using allowed amounts based on Blue Plans' claims data. The cost of care criteria takes into account outliers, patient level risk factors, and geographic variation, before each facility is assessed against a consistent national benchmark.	
4.	Blue Cross Blue Shield of Illinois and advocate health care[49]	All covered services for Advocate health care, a not-for-profit integrate system	System level	Performance for 12 measures	Global budget target	Combination of Quality Hurdle and Cost Hurdle Models.
5.	Blue Cross Blue Shield of Michigan Hospital P4P program[50]	Hospitalized patients with specific index admissions	Hospital	Composite index of collaborative Quality Initiatives, population based, performance, all-cause readmissions	Diagnosis standardized cost-per-case	Unconditional Model. Payments are based on the weighted sum of quality and cost domains
6.	Blue Cross Massachusetts Alternative Quality Contract [51]	All covered services	Alternative Quality Contract provider organizations	32 ambulatory measures, 32 hospital measures. 5 Quality "gates" for each measure, resulting in different bonus	Global budget target	Unconditional Model. High quality is rewarded as a bonus, can equal up to 10% of global budget. [7]

[7] The AQC can be conceptualized two different programs that are not directly connected: a shared savings program and a quality bonus program.

	Name of program	Services evaluated	Level of attribution	Specification of quality	Specification of cost	Approach to combining quality and cost
				payments. Outcome measures are triple weighted. Non-linear function between quality score and payout. [52,53]		
7.	Buyers Health Care Action Group Purchasing Initiative[54,55]	All services	Care systems in Minneapolis/St. Paul	Patient experience and participation in quality improvement initiatives.	Total costs	Side-by-Side Model
8.	Cigna Care Designation[56]	22 categories of specialist services[8]	Physicians and physician groups	5 domains related to National Committee for Quality Assurance (NCQA) Physician Recognition; Group Board Certification; Composite quality index on adherence to 101 Evidence-Based Medicine (EBM) Rules; American Board of Internal Medicine Process Improvement Module Completion; Certified Bariatric Center Affiliated Surgeons.	Costs related to Episode Treatment Group (ETG) methodology	Conditional Model. Providers are compared by specialty within markets.
9.	Cigna Collaborative Accountable Care[57]	All covered services	Large primary care or multispecialty practices, integrated delivery system, of physician-hospital organization	Composite measure assessing adherence to evidence based medicine for preventive care, chronic care, and acute care.	Unclear	Quality Hurdle Model
10.	Health Partners Relative Resource Use[58]	Primary care, specialty care, and hospitals	Physicians, physician practices, and hospitals	Separate composite measures for primary care, specialty care, and hospitals. Components of composite	Uses NQF endorsed total cost of care measure. Encompasses all	Side-by-Side Model

[8] Allergy and Immunology, Cardiology, Cardio-Thoracic Surgery, Colon and Rectal Surgery, Dermatology, Ear, Nose and Throat, Endocrinology, Family Practice, Gastroenterology, General Surgery, Hematology and Oncology, Internal Medicine, Nephrology, Neurology, Neurosurgery, Obstetrics and Gynecology, Ophthalmology, Orthopedics and Surgery, Pediatrics, Pulmonary, Rheumatology, and Urology

	Name of program	Services evaluated	Level of attribution	Specification of quality	Specification of cost	Approach to combining quality and cost
				differ for different types of services.	services with/without price standardization.	
11.	Hospital Value-Based Purchasing	Part A and Part B Medicare services	Hospital	Sum of performance score (incorporating attainment and improvement) for individual measures in various domains (outcomes, clinical process, and patient experience)	Episode covering standardized payments from 3 days prior and 30 days following hospitalization.	Unconditional Model
12.	Integrated Healthcare Association Value Based pay-for-performance program	For seven health plans in California.	Physician organizations	31 clinical quality measures, 15 meaningful use measures, 6 patient experience measures, 12 appropriate resource use measures.	Per member per months total cost of care, including physician, hospital, pharmacy and ancillary payments.	Quality Hurdle and Cost Hurdle Models are used together. Shared savings model then adjusts savings by quality performance.
13.	Leapfrog Hospital Rewards Program[59]	Patients hospitalized with AMI, pneumonia, or child birth, or receiving CABG or PCI.	Hospital	Composite score of multiple measures. Uses a two-level weighting approach based on potential of indicator to reduce mortality and the importance of the indicator to the employer.	Inpatient costs	Conditional Model
14.	Maine Health Management Coalition	Adult care, pediatric care, and hospital care	Physicians, physician practices, and hospitals	Composite measure based on Bridges to Excellence / Hospital Compare measures categorized into “low”, “good”, “better”, and “best”	Whether practice is “working to control cost”	Side-by-Side Model
15.	Maryland multi-payer patient-centered medical home program[49]	All covered services	Primary care practices	21 quality measures; and reductions in use of high-cost services.	Total costs for assigned patients.	The Cost Hurdle Model.
16.	Massachusetts Group Insurance Commission value-	All covered services	Physician-level. Physician profiles various participating	Composite based on 79 quality measures relevant to particular providers	Price standardized episode costs based on Symmetry	Unclear

	Name of program	Services evaluated	Level of attribution	Specification of quality	Specification of cost	Approach to combining quality and cost
	tiering program[60,61]		plans		Episode Treatment Group methodology[62]	
17.	Medica and Fairview health services[49]	All covered services for Fairview Health Services, a non-profit health system	System level	Minimum quality gate, then confidential algorithm	Global budget target	Unclear
18.	Medicare Physician Group Practice Demonstration	Part A and Part B Medicare services	Integrated delivery systems	Performance for 32 ambulatory care performance measures.	Total costs per capita for aligned beneficiaries	Unconditional Model (it seems)
19.	Medicare Shared Savings and Pioneer Accountable Care Organization programs[63]	Part A and Part B Medicare services	Accountable Care Organization	Composite measure of patient/caregiver experience (7 measures); Care coordination/patient safety (6 measures); Preventive health (8 measures); At-risk population: Diabetes (1 measure and 1 composite consisting of five measures); Hypertension (1 measure) Ischemic Vascular Disease (2 measures); Heart Failure (1 measure); Coronary Artery Disease (1 composite consisting of 2 measures).	Payment standardized total costs per capita for aligned beneficiaries	Quality Hurdle Model
20.	NCQA relative resource use[64]	Condition-specific costs for people with specified chronic diseases.[9]	Health plan level by product (e.g. HMO, PPO)	Composite measure based on HEDIS indicators relevant to disease area	Annual condition-specific costs for all relevant services	Side-by-Side Model
21.	Physician Value-	Part A and Part B	Physician practice	Composite measure of clinical	Composite measure	Conditional Model

[9] Asthma, cardiovascular conditions, COPD, diabetes, and hypertension

	Name of program	Services evaluated	Level of attribution	Specification of quality	Specification of cost	Approach to combining quality and cost
	Based payment modifier	Medicare services		care, patient experience, population/community health, patient safety, care coordination, and efficiency.	of total costs per capita for attributed beneficiaries, and for beneficiaries with specific chronic disease	
22.	Tufts Health Plan primary care "Blue Ribbon" program[65]	Primary care	Physician practice	7 HEDIS process of care measures and 7 patient experience measures. Calculate adjusted composite process scores (z-scores), and composite scores for patients experience (z-scores). Scores were then summed and renormalized.	Primary care Episode Treatment Groups	The Conditional Model. The quality and cost domains are standardized and combined with equal weighting. To be designated with the "Blue Ribbon", providers must be above the median on both the quality and cost domains, as well as the combined domain.
23.	UnitedHealth Premium[66-68]	25 categories of specialist services.[10]	Physician, physician practices	Composite score based on evidence based measures related to preventive care, appropriate care, chronic disease care, patient safety, sequencing of care, and care outcomes.	Risk adjusted total cost of care (population cost), and episode cost measurement.	The Unconditional Model. Provider designations are made separately for cost and quality based on statistical criteria. It's unclear how the different designations translate into payment or cost sharing differences.

[10] Allergy, Cardiology, Cardiology – Electrophysiology, Cardiology – Interventional, Endocrinology, Family Medicine, General Surgery, General Surgery - Colon/Rectal, Internal Medicine, Nephrology, Neurology, Neurosurgery – Spine, Ophthalmology, Obstetrics and Gynecology, Orthopedics - Foot/Ankle, Orthopedics – General, Orthopedics – Hand, Orthopedics - Hip/Knee, Orthopedics - Shoulder/Elbow, Orthopedics – Spine, Orthopedics – Sports Medicine, Pediatrics, Pulmonology, Rheumatology, and Urology

	Name of program	Services evaluated	Level of attribution	Specification of quality	Specification of cost	Approach to combining quality and cost
24.	Virginia Cardiac Surgery Quality Initiative[69]	All cardiac surgical patients	Surgeon and hospital	Extensive structure (volume), process, and outcome (mortality and complication) measures.	Normalized hospital and surgeon charges[70]	Side-by-side Model. Comparisons are made for anonymized hospitals and are primarily on quality measures.

627

628

Exhibit 2. Distribution of quality and cost measures used in models

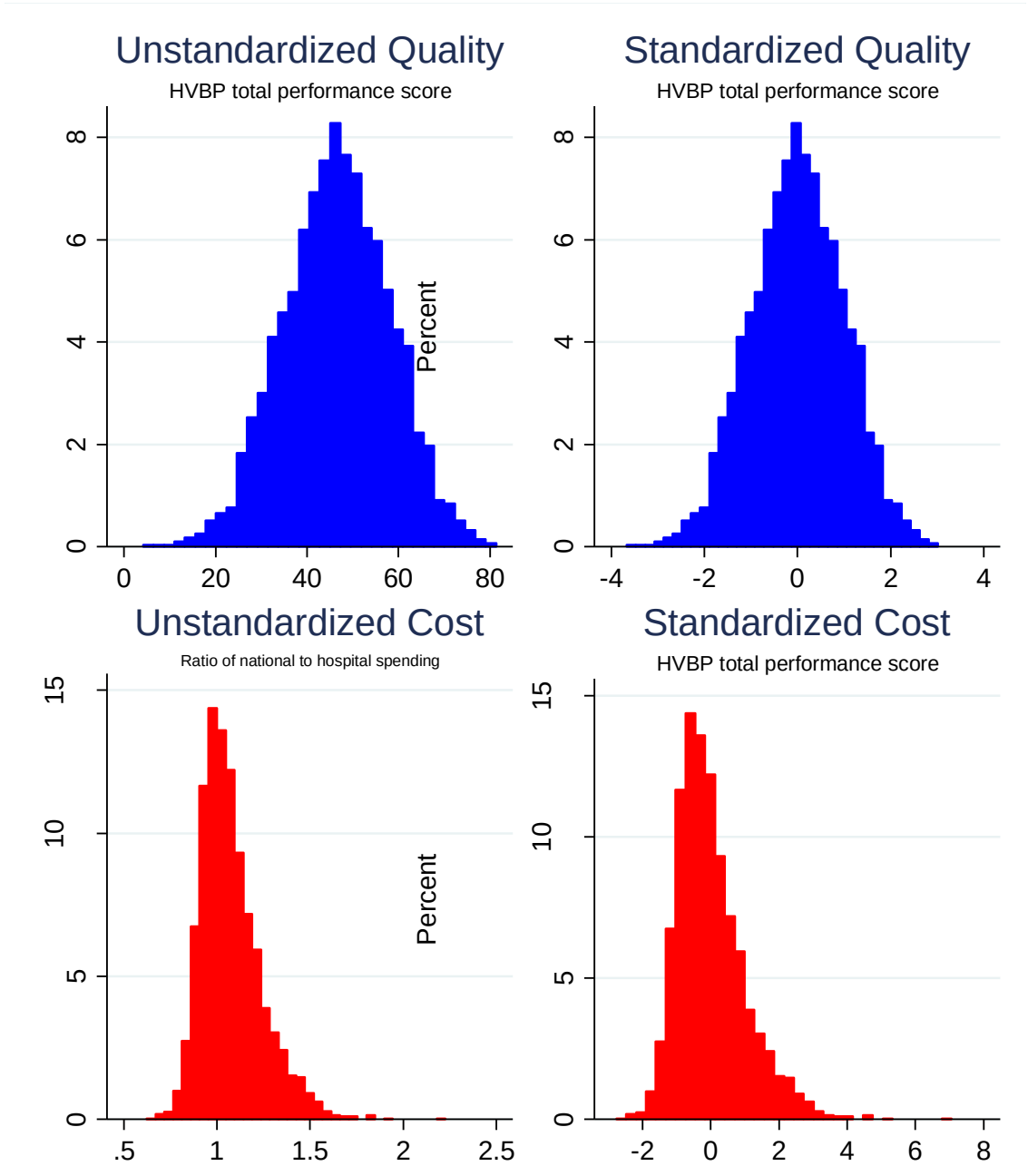
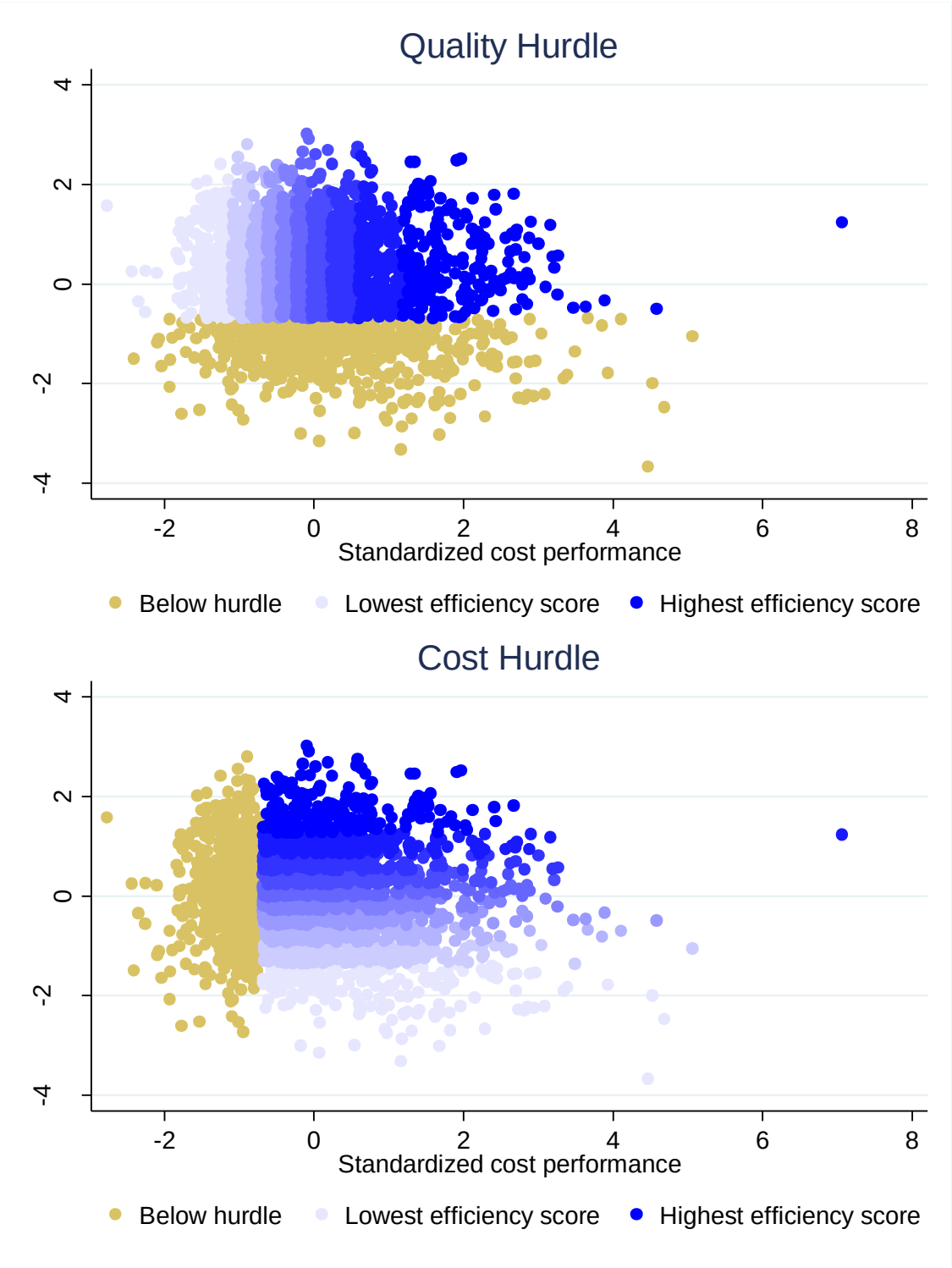


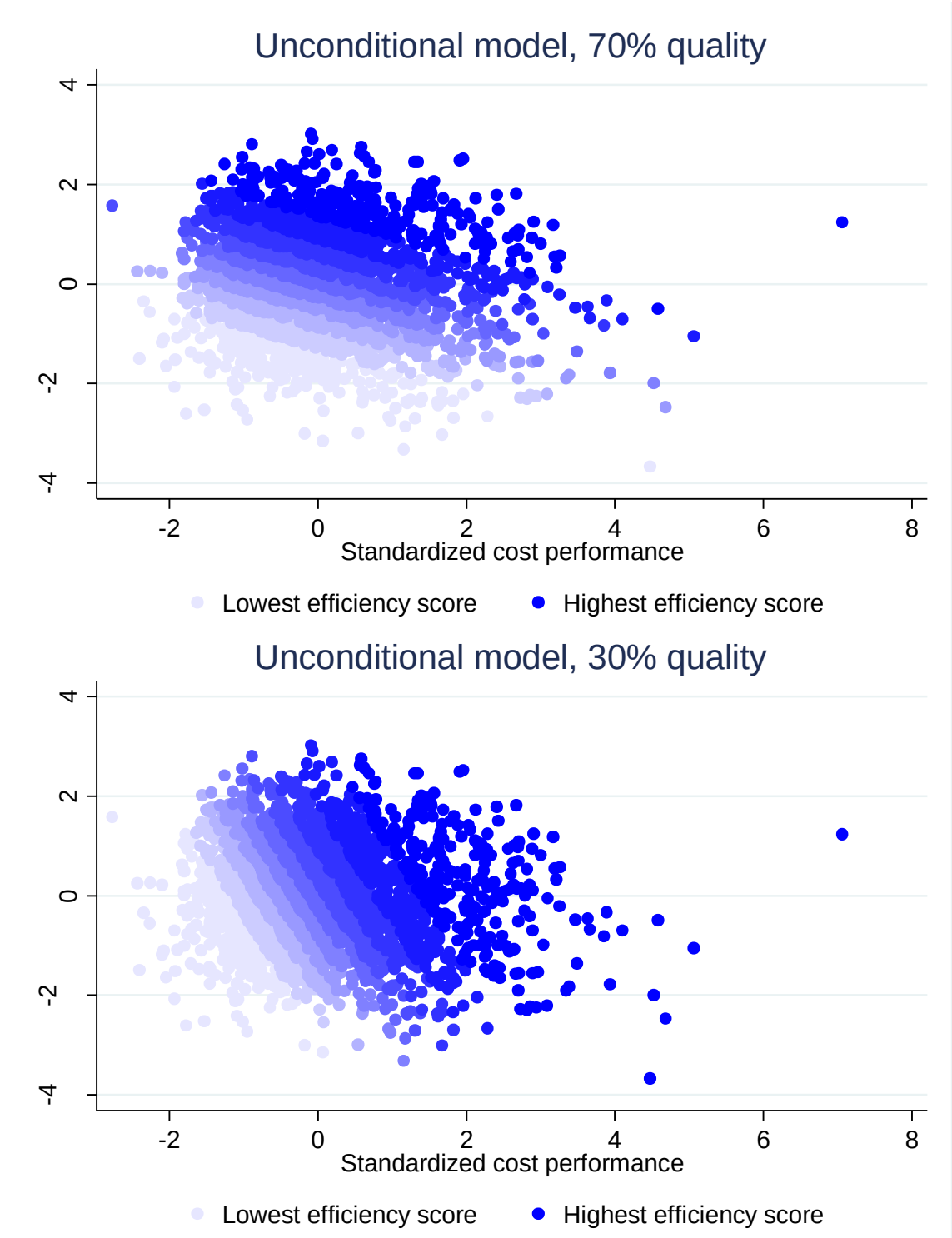
Exhibit 3. Illustration of quality hurdle and cost hurdle models



635

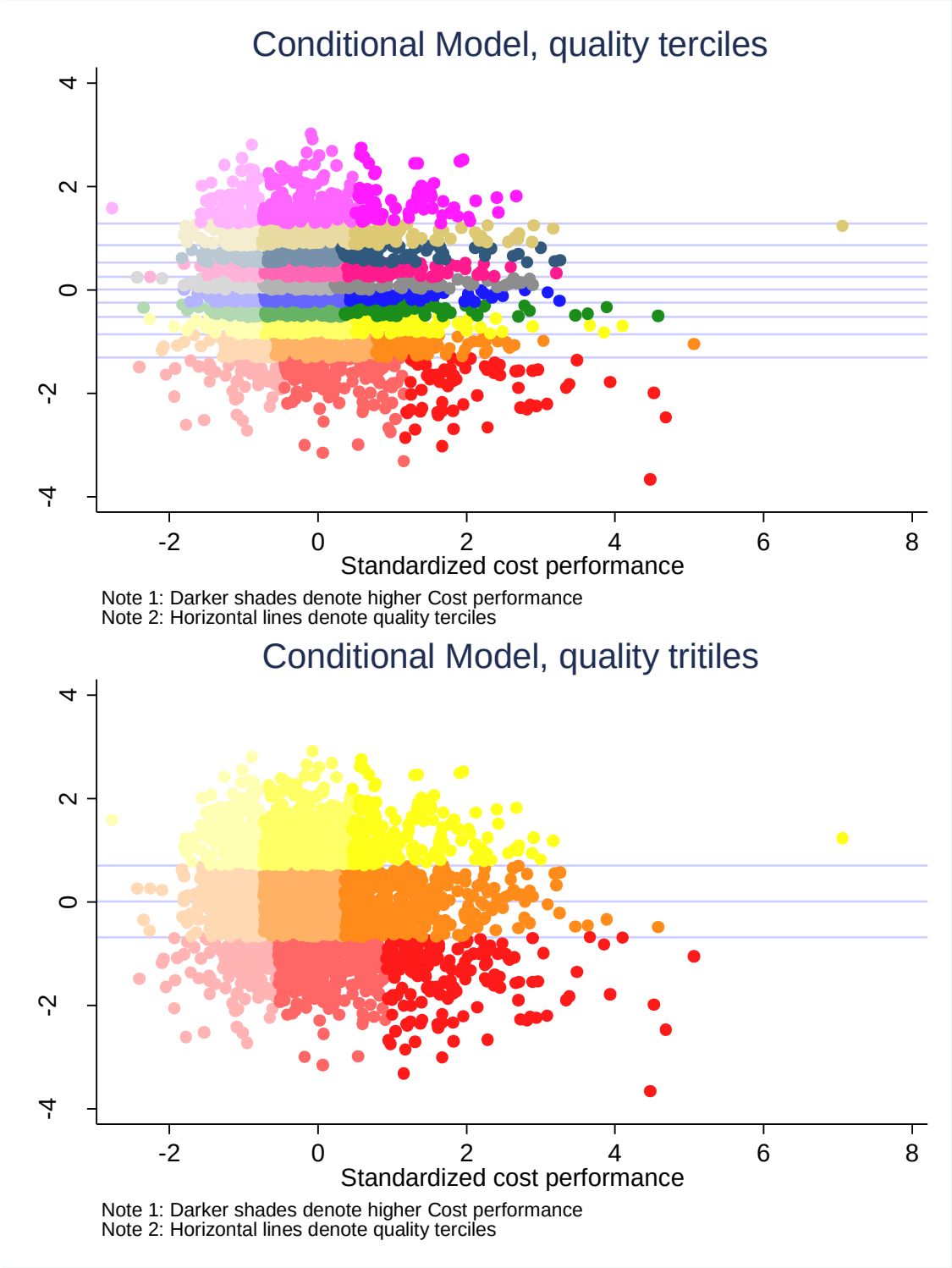
636

637 **Exhibit 4**. Illustration of unconditional model


Unconditional model, 70% quality
4
2
0
-2
-4
-2
0
2
4
6
8
Standardized cost performance
Lowest efficiency score
Highest efficiency score
Unconditional model, 30% quality
4
2
0
-2
-4
-2
0
2
4
6
8
Standardized cost performance
Lowest efficiency score
Highest efficiency score

638

639 **Exhibit 5**. Illustration of conditional model



640

641

Exhibit 6. Illustration of stochastic frontier model

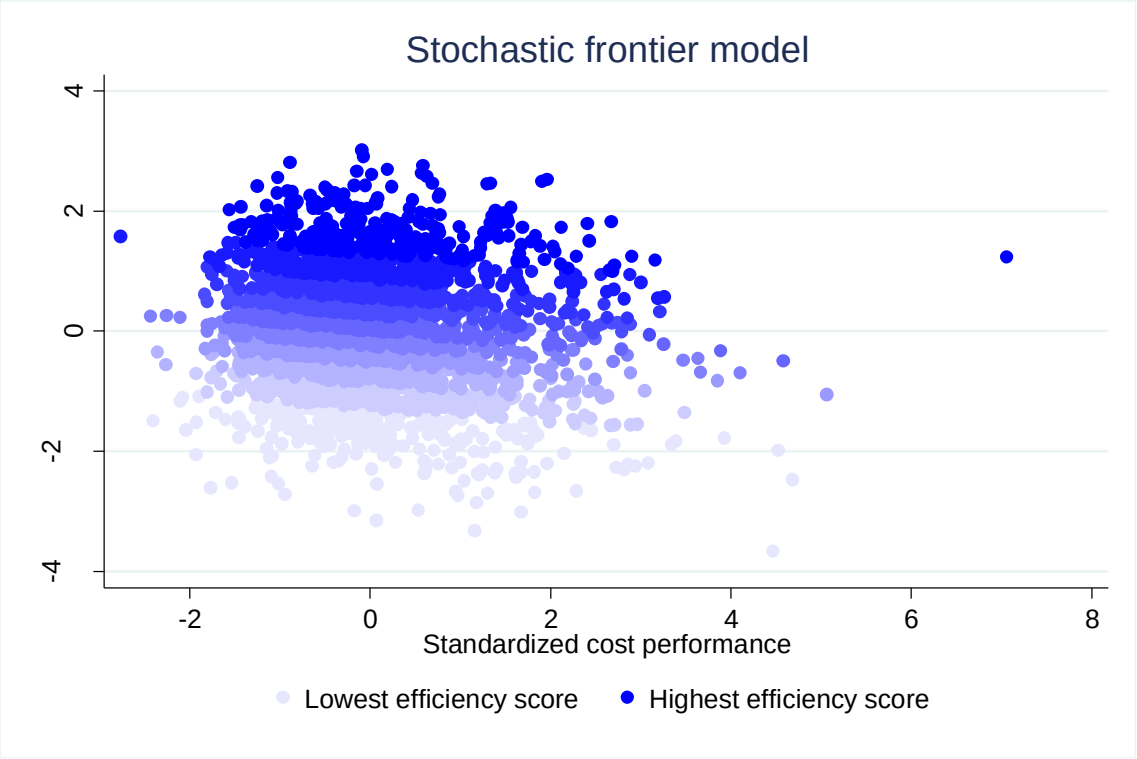


Exhibit 7. Correlation between efficiency scores generated from alternative models linking cost and quality indicators

Model	Quality hurdle	Cost hurdle	Unconditional (70% quality)	Unconditional (30% quality)	Conditional
Quality hurdle	-	-	-	-	-
Cost hurdle	0.1003	-	-	-	-
Unconditional (70% quality)	0.3196	0.8055	-	-	-
Unconditional (30% quality)	0.7802	0.2590	0.6610	-	-
Conditional	0.2122	0.7591	0.8906	0.5118	-
Frontier	0.0718	0.8745	0.9492	0.3992	0.8753

Exhibit 8. Hypothetical example of cost per beneficiary and total quality scores for sample hospitals (positive correlation)

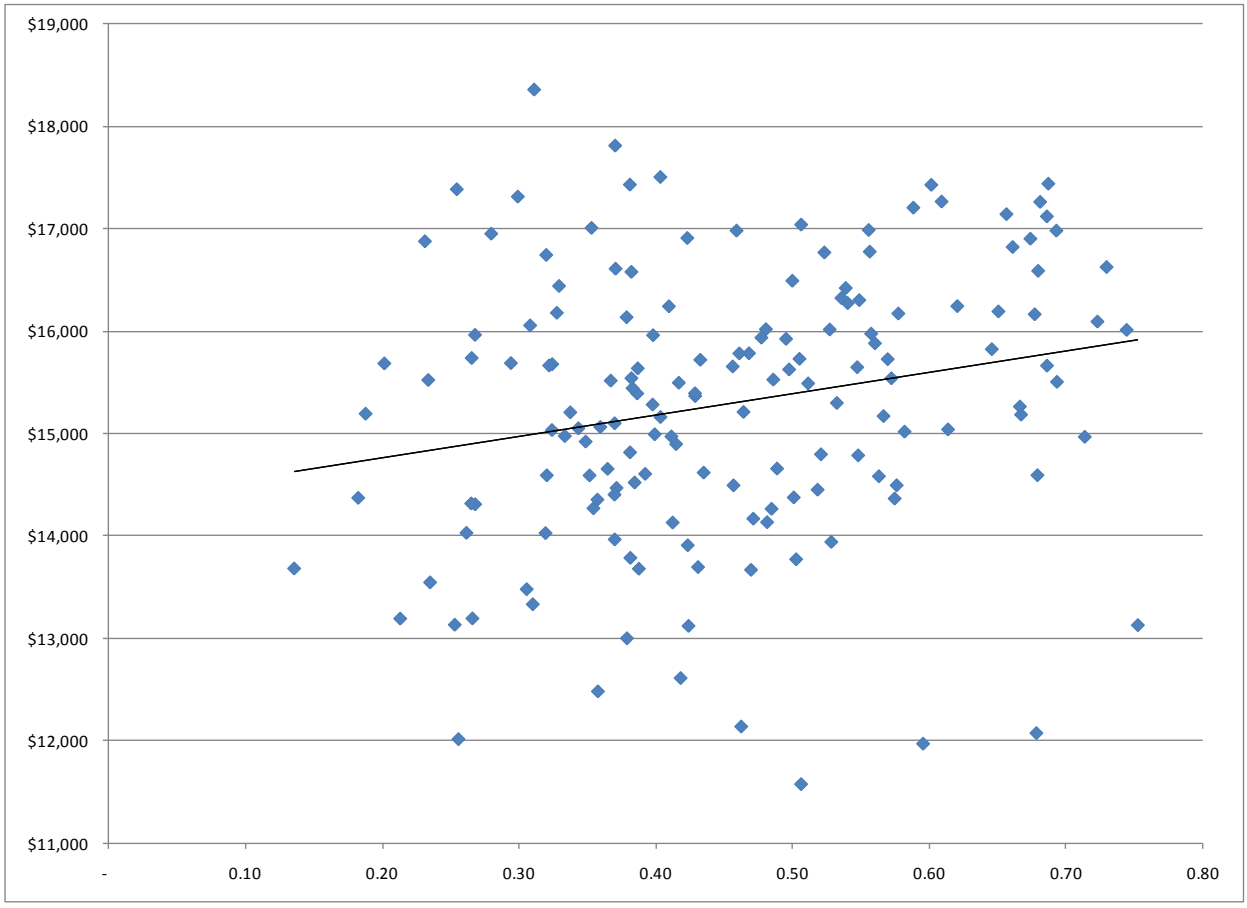


Exhibit 9. Hypothetical example of cost per Beneficiary and quality scores for modified sample (negative correlation)

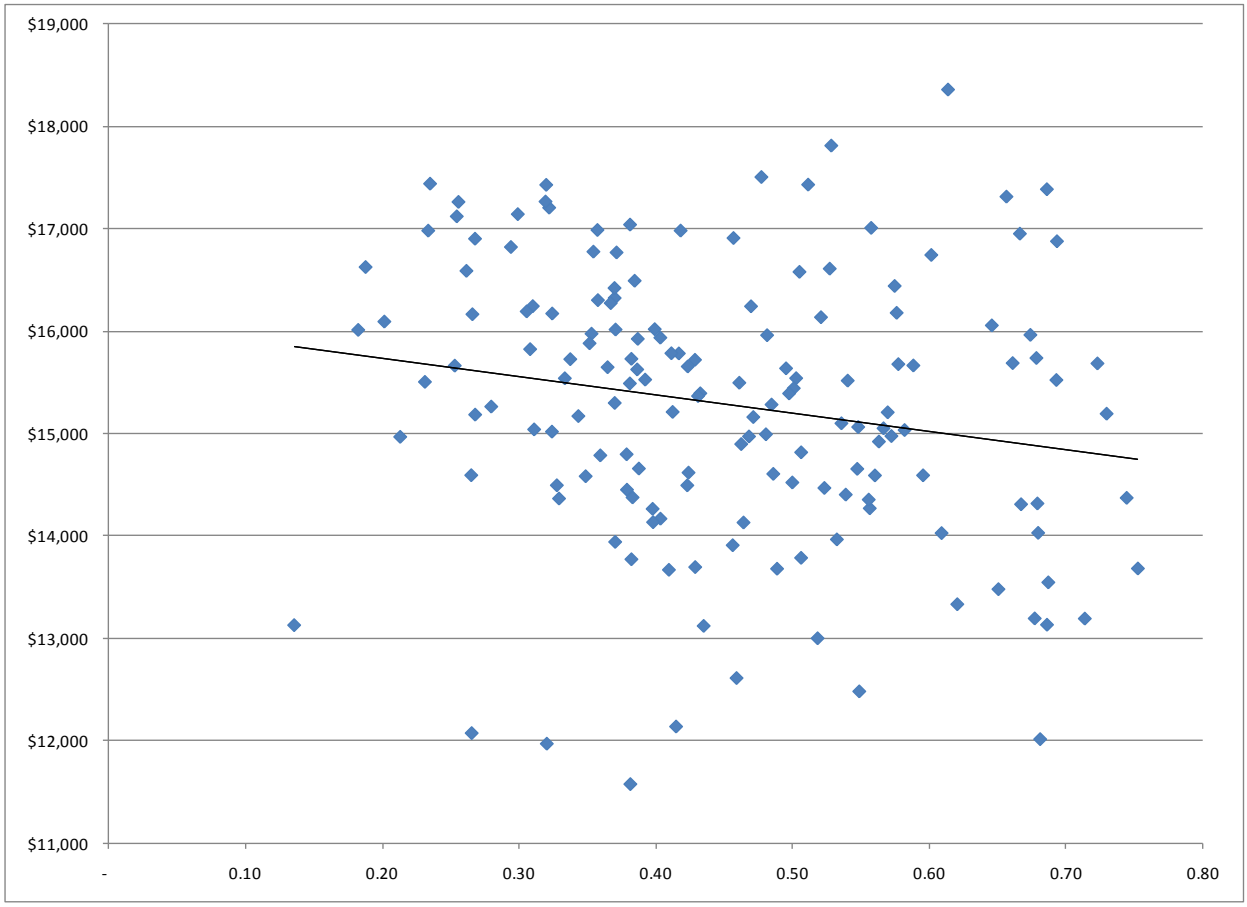


Exhibit 10. Illustrative efficiency value system

	Higher Eff	↑	10	**	**	***	***	****	****	*****	*****	*****	*****
			9	*	**	***	***	****	****	****	*****	*****	*****
			8	*	**	**	***	***	****	****	*****	*****	*****
			7	*	**	**	**	***	***	****	****	*****	*****
Efficiency			6	*	*	**	**	***	***	****	****	*****	*****
Score			5	*	*	**	**	***	***	****	****	*****	*****
			4	*	*	**	**	***	***	****	****	****	*****
			3	*	*	*	**	**	***	***	****	****	*****
			2	*	*	*	**	**	**	***	***	****	*****
			1	*	*	*	*	**	**	***	***	****	****
				1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
				→									
				Higher Quality									
				Total Quality Score									

Legend

***** = **1 Star Efficiency Rating (Lowest Rating)**

****** = **2 Star Efficiency Rating**

******* = **3 Star Efficiency Rating**

******** = **4 Star Efficiency Rating**

********* = **5 Star Efficiency Rating (Highest Rating)**

Exhibit 11. Adjusting the total quality score for efficiency [11]

	10	-1 pt	-1 pt	0 pts	0 pts	+ 1 pt	+ 1 pt	+2 pts	+2 pts	+2 pts	+2 pts
	9	-2 pts	-1 pt	0 pts	0 pts	+ 1 pt	+ 1 pt	+ 1 pt	+2 pts	+2 pts	+2 pts
	8	-2 pts	-1 pt	-1 pt	0 pts	0 pts	+ 1 pt	+ 1 pt	+2 pts	+2 pts	+2 pts
	7	-2 pts	-1 pt	-1 pt	-1 pt	0 pts	0 pts	+ 1 pt	+ 1 pt	+2 pts	+2 pts
Efficiency	**6**	-2 pts	-2 pts	-1 pt	-1 pt	0 pts	0 pts	+ 1 pt	+ 1 pt	+2 pts	+2 pts
Score	**5**	-2 pts	-2 pts	-1 pt	-1 pt	0 pts	0 pts	+ 1 pt	+ 1 pt	+2 pts	+2 pts
	4	-2 pts	-2 pts	-1 pt	-1 pt	0 pts	0 pts	+ 1 pt	+ 1 pt	+ 1 pt	+2 pts
	3	-2 pts	-2 pts	-2 pts	-1 pt	-1 pt	0 pts	0 pts	+ 1 pt	+ 1 pt	+2 pts
	2	-2 pts	-2 pts	-2 pts	-1 pt	-1 pt	-1 pt	0 pts	0 pts	+ 1 pt	+2 pts
	1	-2 pts	-2 pts	-2 pts	-2 pts	-1 pt	-1 pt	0 pts	0 pts	+ 1 pt	+ 1 pt
		1-10	**11-20**	**21-30**	**31-40**	**41-50**	**51-60**	**61-70**	**71-80**	**81-90**	**91-100**
		Total Quality Score									

Summary Table

*** = 1 Star Efficiency Rating (Lowest Rating) = Total Quality Score – 2 pts**

**** = 2 Star Efficiency Rating = Quality Points – 1 pt**

***** = 3 Star Efficiency Rating = Quality Points + 0 pts**

****** = 4 Star Efficiency Rating = Quality Points + 1 pt**

******* = 5 Star Efficiency Rating (Highest Rating) = Total Quality Score + 2 pts**

[11] It is also possible to construct a function that approximates the point assignments in Exhibit 11 (and by extension the star assignment in Exhibit 10) rather than make individual determinations of the point values that should be assigned to each cell. For example, the grid in Exhibit 10 represents a function where quality is weighted twice as much as efficiency in determining the point totals. The function is of the form: a*quality score + b*efficiency score + k ,with a max/min of +/- 2. If the max/min of +/-2 is removed, the highest point adjustments would be +/- 3 rather than +/- 2.

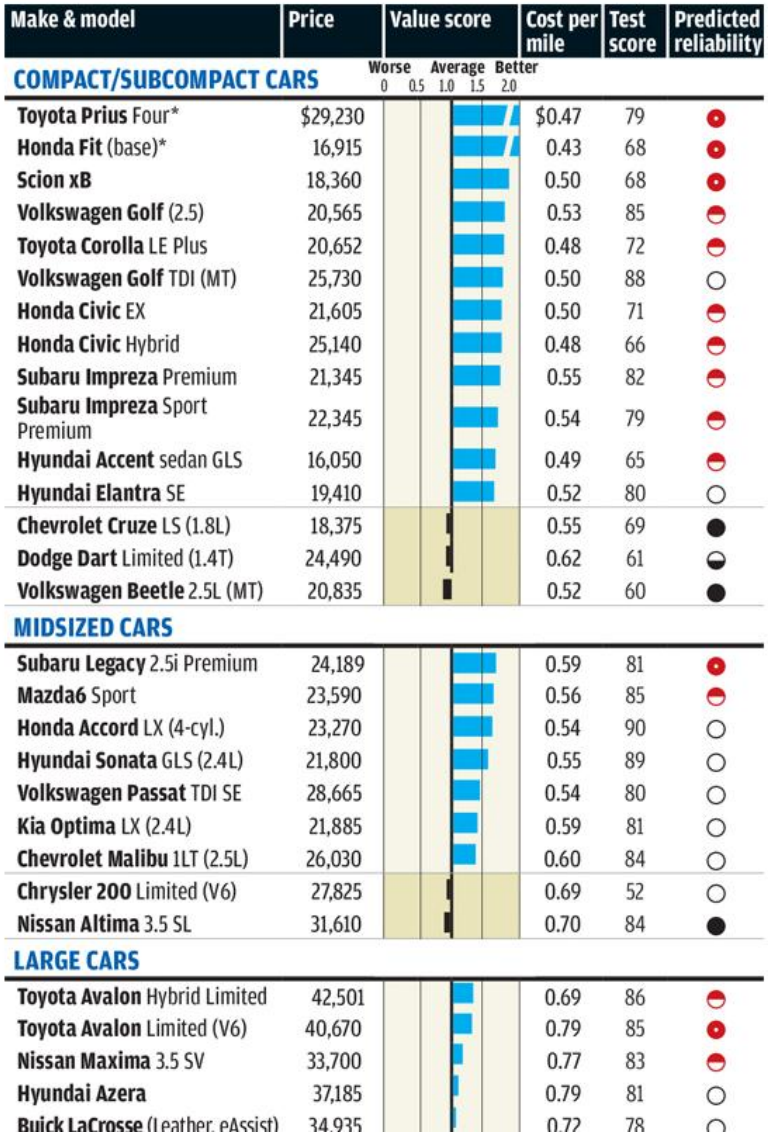
Appendix A: Applications outside of health care that combine indicators of quality and cost

There are a number of efforts outside of health care that link measures of products' cost and quality to measure "value." These include efforts by Consumer Reports to measure the value of automobiles and US News and Weekly Report's rankings of the value of colleges. Consumer Reports' 2014 rankings of the "Best and Worst Cars for the Money" assigns cars a "value-score" based on the combination of a "road test score" (i.e., quality), a composite rated on a 100 point scale, a predicted reliability score (assessing how well the car is likely to hold up given the reliability of recent models), and 5-year ownership costs. This value score is then displayed alongside the car's price and the component factors comprising the value score (ownership costs, road test score, and predicted reliability). The rankings also use a fudge factor to "not recommend" cars as a result of poor reliability. They are able to do this because the ratings are concerned primarily with identifying highly recommended cars: it is therefore not concerned about precise measurements of poor value, and instead focuses on precisely measuring the highest value cars. Value rankings are then displayed within classes of cars (e.g. compact cars, midsized cars, luxury cars, small sport utility vehicles, midsized sport utility vehicles, etc.). See **Appendix Exhibit 1** for how these data are displayed. The model used to calculate value rankings is not specified. Consumer Reports' automobile value ratings have a close analogue to efficiency-tiering in health care. These programs are frequently designed to direct patients towards higher efficiency providers (through a "designation" program) rather than directing patients away from lower efficiency providers and report efficiency for different "classes" of physicians (i.e., different specialties).

US News and World Reports compiles a list of the highest value colleges and universities in their "Best Value Schools."[71] To do this, they first assess school quality as a composite of the student selectivity, college graduation rates, assessment from peer institutions, faculty resources (i.e., class size), financial resources, and alumni giving. They then create a value score as the weighted combination of three factors: 1) 60% is for the ratio of quality to cost (including tuition, room and board, books, and other expenses), discounted according to the average

need-based scholarship; 2) an unspecified percentage is based on the percentage of students who received need-based aid; 3) an unspecified percentage is based the percentage of total costs that are discounted. Similar to Consumer Reports' automobile rankings, US News ranks the value of schools within different categories (e.g. national universities, regional universities, national liberal arts colleges, and regional liberal arts colleges). When showing the rankings of schools based on value, US News and World Reports displays some of the component parts side-by-side (percentage of students receiving need-based grants and the average cost after receiving grants based on need), but not all of the parameters going into the value calculation (see **Appendix Exhibit 2**). This model for combining quality and cost to measure value is similar to the unconditional model with a large weight towards cost.

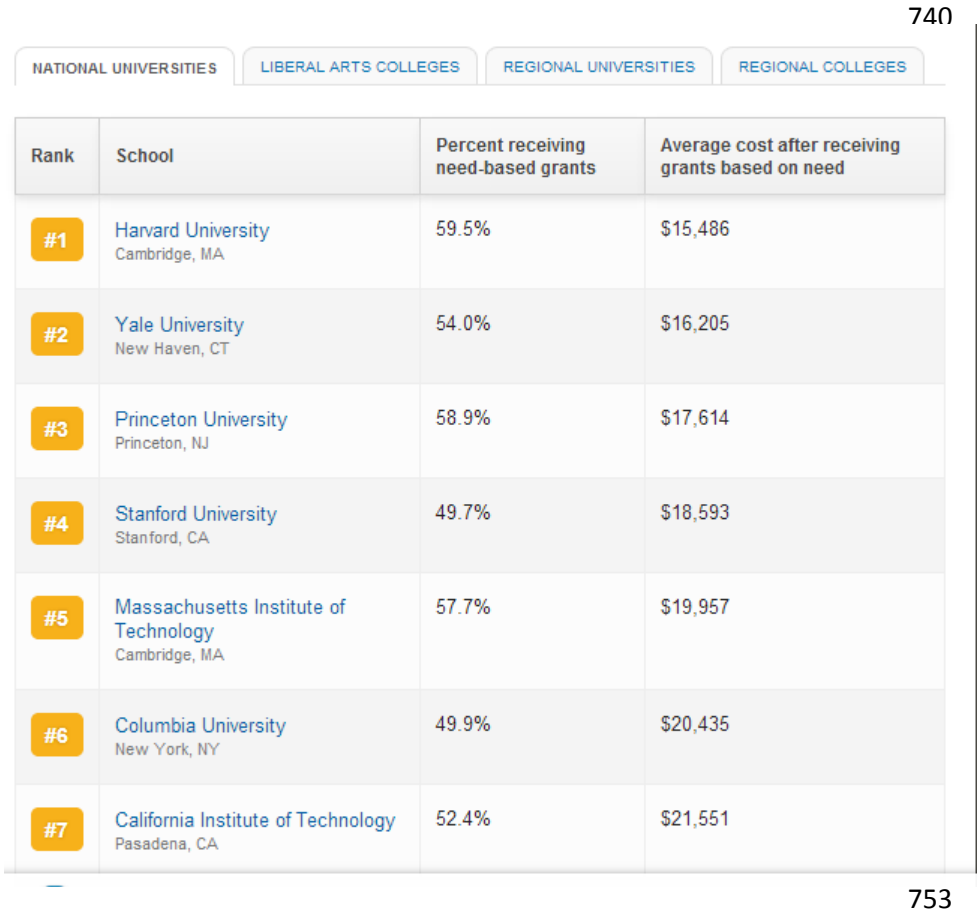
Appendix Exhibit 1. Screen shot from Consumer Reports 2014 "Best and Worst Cars for the Money"



Make & model	Price	Value score (Worse 0 0.5 Average 1.0 1.5 Better 2.0)	Cost per mile	Test score	Predicted reliability
COMPACT/SUBCOMPACT CARS					
Toyota Prius Four*	$29,230		$0.47	79	
Honda Fit (base)*	16,915		0.43	68	
Scion xB	18,360		0.50	68	
Volkswagen Golf (2.5)	20,565		0.53	85	
Toyota Corolla LE Plus	20,652		0.48	72	
Volkswagen Golf TDI (MT)	25,730		0.50	88	
Honda Civic EX	21,605		0.50	71	
Honda Civic Hybrid	25,140		0.48	66	
Subaru Impreza Premium	21,345		0.55	82	
Subaru Impreza Sport Premium	22,345		0.54	79	
Hyundai Accent sedan GLS	16,050		0.49	65	
Hyundai Elantra SE	19,410		0.52	80	
Chevrolet Cruze LS (1.8L)	18,375		0.55	69	
Dodge Dart Limited (1.4T)	24,490		0.62	61	
Volkswagen Beetle 2.5L (MT)	20,835		0.52	60	
MIDSIZED CARS					
Subaru Legacy 2.5i Premium	24,189		0.59	81	
Mazda6 Sport	23,590		0.56	85	
Honda Accord LX (4-cyl.)	23,270		0.54	90	
Hyundai Sonata GLS (2.4L)	21,800		0.55	89	
Volkswagen Passat TDI SE	28,665		0.54	80	
Kia Optima LX (2.4L)	21,885		0.59	81	
Chevrolet Malibu 1LT (2.5L)	26,030		0.60	84	
Chrysler 200 Limited (V6)	27,825		0.69	52	
Nissan Altima 3.5 SL	31,610		0.70	84	
LARGE CARS					
Toyota Avalon Hybrid Limited	42,501		0.69	86	
Toyota Avalon Limited (V6)	40,670		0.79	85	
Nissan Maxima 3.5 SV	33,700		0.77	83	
Hyundai Azera	37,185		0.79	81	
Buick LaCrosse (Leather, eAssist)	34,935		0.72	78	

Source: http://www.consumerreports.org/cro/2012/05/best-new-car-values/index.htm

Appendix Exhibit 2. Screen shot from US News and World Reports 2014 "Best Value Schools"



NATIONAL UNIVERSITIES | LIBERAL ARTS COLLEGES | REGIONAL UNIVERSITIES | REGIONAL COLLEGES

Rank	School	Percent receiving need-based grants	Average cost after receiving grants based on need
#1	Harvard University Cambridge, MA	59.5%	$15,486
#2	Yale University New Haven, CT	54.0%	$16,205
#3	Princeton University Princeton, NJ	58.9%	$17,614
#4	Stanford University Stanford, CA	49.7%	$18,593
#5	Massachusetts Institute of Technology Cambridge, MA	57.7%	$19,957
#6	Columbia University New York, NY	49.9%	$20,435
#7	California Institute of Technology Pasadena, CA	52.4%	$21,551

Source: http://colleges.usnews.rankingsandreviews.com/best-colleges/rankings/national-universities/best-value/spp+50

References

1. Crawford S. Linking cost and quality expert panel orientation. Paper presented at: National Quality Forum 2014.
2. Cutler DM. Where are the health care entrepreneurs? The failure of organizational innovation in health care. In: Lerner J, Stern S, eds. *Innovation Policy and the Economy.* Vol 11. Chicago: University of Chicago Press; 2011:1-28.
3. The watch on the job. *New England Journal of Medicine.* 1969;281(14):792-793.
4. Thornton TN, Leonard RC. Experimental comparison of effectiveness and efficiency of three nursing approaches. *Nursing Research.* 1964;13:122-125.
5. Hussey PS, de Vries H, Romley J, et al. A systematic review of health care efficiency measures. *Health Services Research.* 2009;44(3):784-805.
6. National Quality Measures Clearinghouse. Varieties of measures in NQMC. *Tutorials on Quality Measures* http://www.qualitymeasures.ahrq.gov/tutorial/varieties.aspx. Accessed 04/14/2014.
7. Korenstein D, Falk R, Howell E, Bishop T, Keyhani S. Overuse of health care services in the United States: An understudied problem. *Archives of Internal Medicine.* 2012;172(2):171-178.
8. Burke LA, Ryan AM. The complex relationship between cost and quality in US health care. *The Virtual Mentor.* 2014;16(2):124-130.
9. Doyle JJ, Graves JA, Gruber J, Kleiner S. Do high-cost hospitals deliver better care? Evidence from ambulance referral patterns. www.nber.org March 2012.
10. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending. Part 1: the content, quality, and accessibility of care. *Annals of Internal Medicine.* 2003;138(4):273-287.
11. Silber JH, Kaestner R, Even-Shoshan O, Wang Y, Bressler LJ. Aggressive treatment style and surgical outcomes. *Health Services Research.* 2010;45(6 Pt 2):1872-1892.
12. Gao J, Moran E, Almenoff PL, Render ML, Campbell J, Jha AK. Variations in efficiency and the relationship to quality of care in the veterans health system. *Health Affairs (Millwood).* 2011;30(4):655-663.
13. National Quality Forum. *Measurement framework: Evaluating efficiency across patient-focused episodes of care.* Washington D.C.: National Quality Forum;2009.
14. Institute of Medicine. *Crossing the quality chasm: The IOM health care quality initiative.* Washington, D.C.: Institute of Medicine;2001.
15. Leatherman S, Berwick D, Iles D, Lewin L, Davidoff F, Nolan T, Bisognano M. The business case for quality: case studies and an analysis. *Health Affairs (Millwood).* 2003;22(2):17-30.
16. Ryan AM. Effects of the Premier Hospital Quality Incentive Demonstration on Medicare patient mortality and cost. *Health Services Research.* 2009;44(3):821-842.
17. Robinson JC. The end of managed care. *Journal of the American Medical Association.* 2001;285(20):2622-2628.
18. VanLare JM, Blum JD, Conway PH. Linking performance with payment: implementing the Physician Value-Based Payment Modifier. *Journal of the American Medical Association.* 2012;308(20):2089-2090.
19. Ryan A, Blustein J. Making the best of hospital pay for performance. *The New England Journal of Medicine.* 2012;366(17):1557-1559.
20. Ryan AM, Damberg CL. What can the past of pay-for-performance tell us about the future of Value-Based Purchasing in Medicare? *Healthcare.* 2013;1(1-2):42-49.
21. Fisher ES, McClellan MB, Safran DG. Building the path to accountable care. *New England Journal of Medicine.* 2011;365(26):2445-2447.

22. *Variation in Health Care Spending: Target Decision Making, Not Geography.* The National Academies Press; 2013.
23. Robinson JC. Hospital tiers in health insurance: Balancing consumer choice with financial incentives. *Health Affairs (Millwood).* 2003;Suppl Web Exclusives:W3-135-146.
24. Obamacarefacts.com. Types of Health Insurance Plans. 2014; http://obamacarefacts.com/insurance-exchange/health-insurance-plans.php Accessed April 14, 2014, 2014.
25. Wall JK. Amid Obamacare's changes, WellPoint keeps old playbook. *The Dose.* July 25, 2013 www.ibj.com: Indianapolis Business Journal; 2013.
26. Robinson JC, Brown TT. Increases in consumer cost sharing redirect patient volumes and reduce hospital prices for orthopedic surgery. *Health Affairs (Millwood).* 2013;32(8):1392-1397.
27. Sennett C, Starkey K. *Measuring and improving efficiency in Health Care: Report from an ABIM foundation/IOM meeting.* Philadelphia: ABIM Foundation;2006.
28. Krumholz HM, Keenan PS, Brush JE, Jr., et al. Standards for measures used for public reporting of efficiency in health care: A scientific statement from the American Heart Association Interdisciplinary Council on Quality of Care and Outcomes Research and the American College of Cardiology Foundation. *Circulation.* 2008;118(18):1885-1893.
29. Timbie JW, Normand SL. A comparison of methods for combining quality and efficiency performance measures: Profiling the value of hospital care following acute myocardial infarction. *Statistics in Medicine.* 2008;27(9):1351-1370.
30. Tompkins CP, Higgins AR, Ritter GA. Measuring outcomes and efficiency in Medicare Value-Based Purchasing. *Health Affairs (Millwood).* 2009;28(2):w251-261.
31. Kessler D, McClellan M. *The effects of hospital ownership on medical productivity.* Cambridge, MA: National Bureau of Economic Research;2001.
32. Hollingsworth B. The measurement of efficiency and productivity of health care delivery. *Health Economics.* 2008;17(10):1107-1128.
33. Rosko MD. Measuring technical efficiency in health care organizations. *Journal of Medical Systems.* 1990;14(5):307-322.
34. Meltzer DO, Chung JW. The population value of quality indicator reporting: A framework for prioritizing health care performance measures. *Health Affairs (Millwood).* 2014;33(1):132-139.
35. Berenson RA. *Moving payment from volume to value: What role for performance measurement?* Urban Institute;2010.
36. Massachusetts Division of Health Care Finance and Policy. *Massachusetts health care cost trends: Price variation in health care services.* Boston, MA 2011.
37. The Leapfrog Group. Survival predictor (surgical mortality). Washington D.C.: Academy Health; 2011.
38. Hibbard J, Sofaer S. *Best practices in public reporting No. 1: How to effectively present health care performance data to consumers.* Rockville, MD: Agency for Healthcare Research and Quality;May 2010.
39. Krumholz HM, Wang Y, Mattera JA, Wang Y, Han LF, Ingber MJ, Roman S, Normand SL. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation.* 2006;113(13):1683-1692.
40. Burack JH, Impellizzeri P, Homel P, Cunningham JN. Public reporting of surgical mortality: A survey of New York State cardiothoracic surgeons. *Annals of Thoracic Surgery.* 1999;68(4):1195-1200.

41. Silber JH, Rosenbaum PR, Ross RN, Ludwig JM, Wang W, Niknam BA, Mukherjee N, Saynisch PA, Even-Shoshan O, Kelz RR, Fleisher LA. Template matching for auditing hospital cost and quality. *Health Services Research.* 2014;epub ahead of print.
42. National Quality Forum. *Composite performance measure evaluation guidance.* Washington, DC: National Quality Forum;2013.
43. Oregon Select Network Comparison Guide. http://www.or.regence.com/network/. Accessed 04/14/2014, 2014.
44. Phinney S. Regence BlueShield sued for defamation. *Seattle pi:* 2006.
45. Aetna. *Aexcel performance network designation measurement methodology.* Aetna;2014.
46. Blue Cross Blue Shield Association. Blue Distinction Centers: An overview. http://www.bcbs.com/why-bcbs/blue-distinction/. Accessed 04/24/2014.
47. Blue Cross Blue Shield Association. Blue Distinction Specialty Care Program: 2013 Program selection criteria for Blue Distinction Centers for cardiac care. http://www.bcbs.com/healthcare-partners/blue-distinction-for-providers/cardiacprogramcriteria.pdf. Accessed 04/24/2014.
48. Blue Cross Blue Shield Association. Blue Distinction Specialty Care Program: 2011-2012 Program selection criteria for Blue Distinction Centers for spine surgery and Blue Distinction Centers for knee and hip replacement. http://www.bcbs.com/healthcare-partners/blue-distinction-for-providers/BDC_SpineKneeHip_Selection_Criteria.pdf. Accessed 04/24/2014.
49. Bailit M, Hughes C, Burns M, Freedman DH. *Share-savings payment arrangements in health care: Six case studies.* New York, NY: The Commonwealth Fund;2012.
50. Blue Cross Blue Shield of Michigan. *2014 Hospital pay-for-performance program: Peer groups 1-4.* Michigan2014.
51. *Blue Cross Blue Shield of Massachusetts: The Alternative Quality Contract.* Blue Cross Blue Shield of Massachusettes; May 2010 2010.
52. Chernew ME, Mechanic RE, Landon BE, Safran DG. Private-payer innovation in Massachusetts: The 'Alternative Quality Contract'. *Health Affairs (Millwood).* 2011;30(1):51-61.
53. Song Z, Safran DG, Landon BE, He Y, Mechanic RE, Day MP, Chernew ME. The 'Alternative Quality Contract,' based on a global budget, lowered medical spending and improved quality-Supplemental Materials. *Health Affairs (Millwood).* 2012;31(8):1885-1894.
54. Christianson JB, Feldman R. Evolution in the Buyers Health Care Action Group purchasing initiative. *Health Affairs (Millwood).* 2002;21(1):76-88.
55. Lyles A, Weiner JP, Shore AD, Christianson J, Solberg LI, Drury P. Cost and quality trends in direct contracting arrangements. *Health Affairs (Millwood).* 2002;21(1):89-102.
56. Cigna. *Cigna care designation and physician quality and cost-efficiency displays: 2013 Methodologies Whitepaper.* Cigna;2013.
57. Cigna. *Collaborative accountable care: CIGNA's approach to Accountable Care Organizations.* Cigna;2011.
58. HealthPartners. Cost and quality ratings:Medical group and hospital ratings. 2014; https://www.healthpartners.com/public/cost-and-quality/index.html. Accessed April 14, 2014.
59. The Agency for Healthcare Research and Quality, The Employer Health Care Alliance Cooperative. Efficiency in health care: What does it mean? How is it measured? How can it be used for Value-Based Purchasing? 2006; Madison, WI.
60. Carroll J. Early Tiered networks encounter many obstacles. *Managed Care*: MediMedia Managed Markets; 2007.
61. Alteras T, Silow-Carroll S. *Value-driven health care purchasing: Case study of the Massachusetts Group Insurance Commission.* The Commonwealth Fund;2007.

62. Sinaiko AD, Rosenthal MB. The impact of tiered physician networks on patient choices. *Health Services Research*. 2014: epub ahead of print.
63. RTI International. *Accountable care organization 2013 program analysis: Quality performance standards narrative measure specifications.* Baltimore, MD: Centers for Medicare & Medicaid Services;2012.
64. NCQA. Relative Resource Use. 2014; http://www.ncqa.org/tabid/1231/default.aspx. Accessed 04/14/2014, 2014.
65. *Tufts Health Plan Blue Ribbon Value Methodology- Summary.* Tufts Health Plan;2006.
66. UnitedHealthCare. *UnitedHealth Premium Physician Designation Program: Summary methodology.* UnitedHealthCare;2014.
67. UnitedHealthCare. UnitedHealth Premium Physician Designation Program: Detailed Methodology. 2014.
68. United Healthcare Online. Premium Methodology. https://www.unitedhealthcareonline.com/b2c/CmaAction.do?channelId=45dff1ab39f24210VgnVCM100000b640dd0a____ . Accessed 04/24/2014.
69. Virginia Cardiac Surgery Quality Initiative. 2014; http://www.vcsqi.org.
70. Speir AM, Kasirajan V, Barnett SD, Fonner E. Additive costs of postoperative complications for isolated coronary artery bypass grafting patients in Virginia. *The Annals of Thoracic Surgery.* 2009;88(1):40-46.
71. Morse R. Best value schools methodology. *US News and World Report*2013. http://www.usnews.com/education/best-colleges/articles/2013/09/09/best-value-schools-methodology