Health Care Efficiency

A Systematic Review of Health Care Efficiency Measures

Peter S. Hussey, Han de Vries, John Romley, Margaret C. Wang, Susan S. Chen, Paul G. Shekelle, and Elizabeth A. McGlynn

Objective. To review and characterize existing health care efficiency measures in order to facilitate a common understanding about the adequacy of these methods.

Data Sources. Review of the MedLine and EconLit databases for articles published from 1990 to 2008, as well as search of the "gray" literature for additional measures developed by private organizations.

Study Design. We performed a systematic review for existing efficiency measures. We classified the efficiency measures by perspective, outputs, inputs, methods used, and reporting of scientific soundness.

Principal Findings. We identified 265 measures in the peer-reviewed literature and eight measures in the gray literature, with little overlap between the two sets of measures. Almost all of the measures did not explicitly consider the quality of care. Thus, if quality varies substantially across groups, which is likely in some cases, the measures reflect only the costs of care, not efficiency. Evidence on the measures' scientific soundness was mostly lacking: evidence on reliability or validity was reported for six measures (2.3 percent) and sensitivity analyses were reported for 67 measures (25.3 percent).

Conclusions. Efficiency measures have been subjected to few rigorous evaluations of reliability and validity, and methods of accounting for quality of care in efficiency measurement are not well developed at this time. Use of these measures without greater understanding of these issues is likely to engender resistance from providers and could lead to unintended consequences.

Key Words. Efficiency, provider profiling, performance measurement, systematic review

Rising health care costs are driving increases in health insurance premiums, the erosion of private coverage (Chernew, Cutler, and Keenan 2005), and strains on the fiscal solvency of public insurance programs (Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Insurance Trust Funds 2007). At the same time, evidence suggests that the efficiency of the U.S. health system—defined by the Institute of Medicine (IOM) as "avoiding

waste, including waste of equipment, supplies, ideas, and energy"—is low (IOM 2001; Fuchs 2005; Bush 2007). Costs vary widely across geographic areas, but the differences are not associated with more reliable delivery of evidence-based care or better health outcomes (Fisher et al. 2003a, b). International comparisons are also often used to question the efficiency of the U.S. health system (Davis et al. 2007).

These concerns have created tremendous pressure to measure the efficiency of health care providers and systems so that it can be evaluated and improved (Cassel and Brennan 2007). Health care quality is now regularly measured, reported, and rewarded with incentive payments. Efficiency measurement has lagged behind. Increasingly, however, large employers have been demanding that health plans incorporate efficiency profiles into their products and information packages alongside quality profiles (Milstein 2004). Health plans have been using provider efficiency ratings in network selection, pay-for-performance programs, or to steer patients toward efficient providers through lower copayments and/or public reporting (Iglehart 2002; American Medical Association 2007; Draper, Liebhaber, and Ginsburg 2007).

Efficiency measurement is also likely to be used increasingly in public programs. President Bush issued an Executive Order in 2006 stipulating that federal health care programs promote quality and efficiency and increase the transparency of relevant information for consumers (The White House 2007). The Medicare Payment Advisory Commission (MedPAC) has advocated using efficiency measurement to improve value in the Medicare program (MedPAC 2007). The IOM included efficiency as one of six aims for the 21st-century health system (IOM 2001).

Despite widespread interest in evaluating efficiency, considerable uncertainty exists about whether the methods are sufficiently well developed to be used outside the research laboratory (Milstein and Lee 2007). First, the term *efficiency* is used by different stakeholders to connote various constructs. Second, little is known about the range of methods that exist to measure efficiency and how well available efficiency metrics capture the constructs of

Address correspondence to Peter S. Hussey, Ph.D., RAND, 1200 S. Hayes St., Arlington, VA 22202; e-mail: hussey@rand.org. Han de Vries, M.S., is with RAND, Westbrook Centre, Cambridge, UK. John Romley, Ph.D., Margaret C. Wang, Ph.D., M.P.H., Susan S. Chen, B.A., and Elizabeth A. McGlynn, Ph.D., are with RAND, Santa Monica, CA. John Romley, Ph.D., is with Occidental College, Los Angeles, CA. Susan S. Chen, B.A., is with College of Agriculture and Natural Resources, Michigan State University, East Lansing, MI. Paul G. Shekelle, M.D., Ph.D., is with Southern California Evidence-based Practice Center—RAND, Santa Monica, CA, and Greater Los Angeles Veterans Affairs Healthcare System, Los Angeles, CA.

interest (The Leapfrog Group and Bridges to Excellence 2007). Payers and purchasers have begun to use efficiency measures despite these uncertainties. Proponents of efficiency measurement seek to "learn on the job" and improve measurements through use. Those who are being evaluated on these metrics worry that the lack of conceptual clarity and the limited methodological assessments increase the likelihood that results from the metrics will create distortions in patterns of care seeking and service delivery, adding to distortions related to current payment systems (O'Kane et al. 2008).

To address the lack of clarity in the concepts and methods of efficiency measurement, we conducted a systematic review of existing efficiency measures and characterized them using a typology we developed. Our work was designed to reach a wide variety of stakeholders, each of which faces different pressures and values in the selection and application of efficiency measures. This paper is intended as the first of several steps that are necessary to create a common understanding among these stakeholders about the adequacy of tools to measure efficiency.

METHODS

We searched for potential measures of health care efficiency in the published literature and "gray" literature, interviewed a sample of vendors who have developed measures, and characterized potential measures according to our typology. This article is based on an Evidence Report prepared for the Agency for Healthcare Research and Quality (McGlynn et al. 2008).

Literature Search

We searched the MedLine and EconLit databases for published articles in the English language that appeared in journals between January 1990 and May 2008 and involved human subjects. Search terms included *efficiency*, *inefficiency*, *productivity*, and *economic profiling*. We also performed "reference mining" by searching the bibliographies of retrieved articles looking for additional relevant publications. Members of the project team worked closely with a technical expert panel and librarians to refine the search strategy. These searches were conducted during December 2005 then updated in May 2008.

Gray Literature

Because some efficiency measures might not appear in the published literature, we relied on experts to develop a list of vendors of measurement

products and other organizations that had developed or were considering developing their own efficiency measures. We contacted key people at these organizations to collect the information necessary to describe and compare their efficiency measures to others we abstracted from publications.

Study Eligibility

In order to be eligible for inclusion, a published study had to present empirical information on an efficiency measure, which we defined as the relationship between a specific product of the health care system (also called an output) and the resources used to create that product (also called inputs). By this definition, a provider in the health care system (e.g., hospital, physician) would be efficient if it was able to maximize output for a given set of inputs or to minimize inputs used to produce a given output. The measured inputs and outputs are assumed to be comparable (discussed in more detail below).

Classification of Measures

We created a typology of efficiency measures to characterize the measures abstracted in the systematic review (McGlynn et al. 2008). The typology has three levels:

- *Perspective*: who is evaluating the efficiency of what entity and what is their objective?
- *Outputs*: what type of product is being evaluated?
- *Inputs*: what inputs that are used to produce the output are included in the measure?

The first tier in the typology, perspective, requires an explicit identification of the entity that is evaluating efficiency, the entity that is being evaluated, and the objective or rationale for the assessment. Each of these three elements of perspective is important because different entities have different objectives for considering efficiency, have control over a particular set of resources or inputs, and may seek to deliver or purchase a different set of services. In classifying measures identified through the scan, however, it was generally feasible only to identify the entity being evaluated. Users of measures should be explicit about their purposes in using efficiency metrics.

The second level of the typology identifies the outputs of interest and how they are measured. We distinguish between two types of outputs: health services (e.g., visits, drugs, and admissions) and health outcomes (e.g., pre-

ventable deaths, functional status, and clinical outcomes such as blood pressure or blood sugar control).

A key issue that arises in efficiency measurement is whether the outputs are comparable, particularly on quality (Newhouse 1994), which is known in some clinical situations to vary widely by patient, condition, provider, and geographic area (Schuster, McGlynn, and Brook 1998; Fiscella et al. 2000). Differences could also exist in the content of a single service, the mix of services in a bundle, and the mix of patients seeking or receiving services. An efficiency measure either can be adjusted for differences in the quality and other aspects of the measured outputs, or it can be used under the assumption that all measured outputs are comparable. However, there is some gray area in how precisely outputs should be adjusted for quality, and the methods for incorporating quality into efficiency measurement are not well developed at this time. Health plans typically attempt to identify high-quality and high-efficiency providers by examining measures of both efficiency and quality, although not necessarily for the same episode, patient, condition, etc. For example, diabetes quality measures (e.g., regular glucose monitoring) could be reported in conjunction with the cost of producing an episode of diabetes care. Another approach would be to adjust the outputs of efficiency measures for quality by directly incorporating quality metrics into the specification of the output. For example, comparisons of the efficiency of producing coronary artery bypass graft surgical procedures would give less weight to procedures resulting in complications. This approach poses significant methodological challenges that can raise concerns about its usefulness at this time. Preliminary research has shown that the results of health care efficiency measurement are sensitive to the method used to incorporate the quality of the measured outputs (Timbie and Normand 2007).

In response to this issue, the AQA, a consortium of physician professional groups, insurance plans, and others, has adopted a principle that measures can only be labeled "efficiency of care" if they incorporate a quality metric; those without quality incorporated are labeled "cost of care" measures (AQA 2007). In our review, we did not exclude purported efficiency measures that did not directly incorporate quality metrics. However, the framework allows for health service outputs to be defined with reference to quality criteria; that is, the framework includes either definition of health services. We used a broader definition in order to capture the range of purported efficiency measurement approaches in use in different applications. As discussed below, use of the AQA definition of an "efficiency measure" as an inclusion criterion would have resulted in the rejection of nearly all of the measures we reviewed.

The third tier of the typology identifies the inputs that are used to produce the outputs of interest that are measured. Inputs can be measured as counts by type (e.g., nursing hours, bed days, and days supply of drugs) or they can be monetized (real or standardized dollars assigned to each unit). We refer to these, respectively, as physical inputs or financial inputs. The way in which inputs are measured may influence the way the results are used. Efficiency measures that count physical inputs help to answer questions about whether the output could be produced faster, with fewer people, less time from people, or fewer supplies. Efficiency measures that use financial inputs help to answer questions about whether the output could be produced less expensively— whether the total cost of labor, supplies, and other capital could be reduced through more-efficient use or substitution of less costly inputs.

We also characterized measures according to several aspects of the methodology used: the statistical/mathematical approach, type of data source, time frame, and explanatory variables included. Finally, we examined whether the "scientific soundness" of each measure was reported; specifically, we searched for reports of empirical evidence of measure reliability and/or validity, as well as analyses of sensitivity to alternate specifications.

Data Synthesis

We classified and narratively summarized the efficiency measures by perspective, outputs, inputs, methods used, and scientific soundness.

RESULTS

Measures in the Published Literature

The electronic literature search identified 5,563 titles (Figure 1). Reference mining identified another 118 potentially relevant titles. Of these, 5,022 were rejected as not relevant to our project, leaving 659 total from all sources. Repeat review by the research team excluded an additional 64 titles. Seven titles could not be located even after contracting with Infotrieve, a private service that specializes in locating obscure and foreign scientific publications. A total of 588 articles were retrieved and reviewed.

Screening of retrieved articles/reports resulted in exclusion of 416: 160 that reported efficiency using non-U.S. data sources; 155 due to research topic (research topic was not health care efficiency measurement); 93 that did not report the results of an efficiency measure; and 8 for other reasons. The re-

Figure 1:   Literature Flow
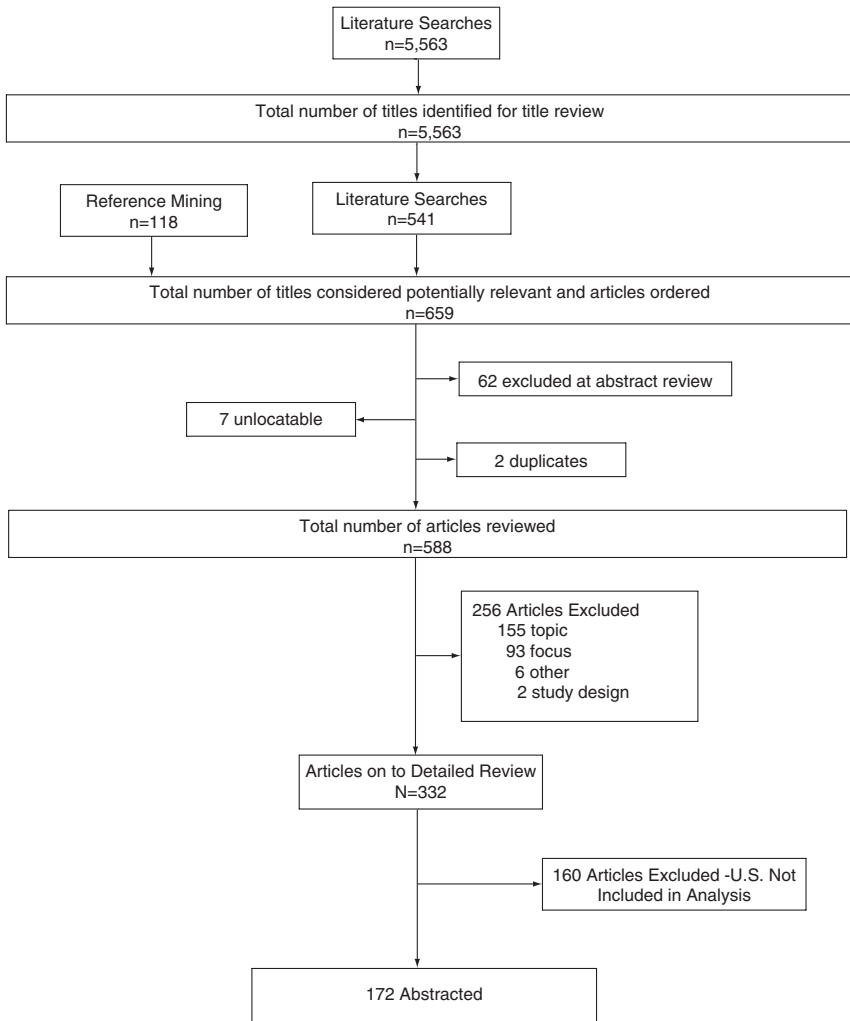
maining 172 articles were accepted for detailed review. These articles contained 265 efficiency measures.

Perspective

The majority of the published literature on health care efficiency has been related to the production of hospital care. Of the 265 efficiency measures

Table 1:   Characteristics of Efficiency Measures Abstracted from Published Literature

	Number of Measures	*Percentage of Measures*
Perspective		
Hospital	162	61.1
Physician (individual or group)	54	20.4
Health plan	13	4.9
Integrated delivery system	5	1.9
Nurse	6	2.3
Geographic region	4	1.5
Medicare program	3	1.1
Other	18	6.8
Inputs		
Physical	123	46.4
Financial	82	30.9
Physical and financial	60	22.6
Outputs		
Health services	258	97.4
Health outcomes	5	1.9
Other	3	1.1
Statistical/mathematical methods		
Frontier analysis or other regression-based approach	147	55.5
Ratios	117	44.2
Explanatory variables*		
Provider level	131	49.4
Patient risk/severity	91	34.3
None	90	34.0
Area level	58	21.9
Patient level (except risk/severity)	38	14.3
Data source		
Secondary data	182	68.7
Primary data	53	20.0
Not available or not applicable	30	11.3
Time frame		
Cross sectional	150	56.6
Longitudinal/time series	104	39.2
Not available	11	4.2
Reporting of scientific soundness		
Reliability and/or validity testing reported	6	2.3
Sensitivity analysis reported	67	25.3

Source: Authors' analysis.

*Some measures included in multiple categories.

abstracted, 162 (61.1 percent) measured the efficiency of hospitals (Table 1). Examples of hospital measures include risk-adjusted average length of stay (Weingarten et al. 2002); cost per risk-adjusted discharge (Conrad et al. 1996);

and the cost of producing both risk-adjusted hospital discharges and hospital outpatient visits (Rosko 2004). Studies of physician efficiency were second most common (54 measures, 20.4 percent). Examples of physician measures in the published literature included relative value units (RVUs) for services provided per physician per month (Conrad et al. 2002); patient visits per physician per month (Albritton et al. 1997); counts of the amounts of resources (labor, equipment, supplies, etc.) used for a medical procedure (Rosenman and Friesner, 2004); and cost per episode of care (Thomas 2006). Much smaller numbers of articles focused on the efficiency of nurses, health plans, other providers, or other entities. None of the abstracted articles reported the efficiency of health care at the national level, although two articles (Ashby, Guterman, and Greene 2000; Skinner, Fisher, and Wennberg 2007) focused on efficiency in the Medicare program.
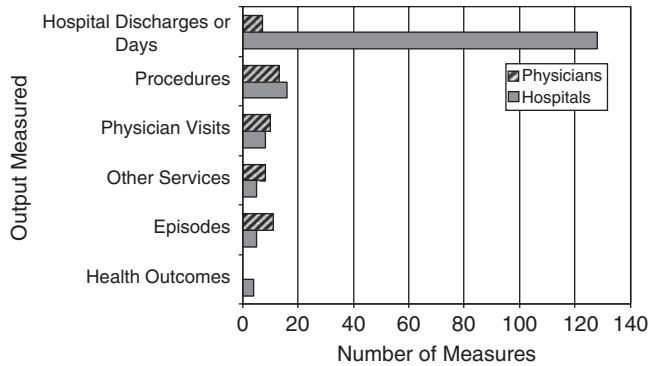
Outputs

Almost all (97.4 percent) of the measures abstracted from the published literature used health services as outputs (Figure 2). Common health service types used as outputs included hospital discharges, procedures, and physician visits. Five measures (Dewar and Lambrinos 2000; Kessler and McClellan 2002; Carey 2003; Skinner, Fisher, and Wennberg 2007) included health outcomes as outputs. One of the measures using health services as outputs explicitly integrated the quality of service provided in the output specification (Harrison and Coppola 2007). A small subset of measures attempted to account for quality by including it as an explanatory variable in a regression model in which efficiency is the dependent variable (Zuckerman, Hadley, and Iezzoni 1994). Some articles also reported results of analyses of quality or outcomes separately from analyses of efficiency.

Inputs

Of the 265 efficiency measures in the published literature, 123 used counts of physical resources as inputs, 82 used the costs of inputs, and 60 used both physical resources and costs (Table 1). Many of the inputs used in measures were used as outputs in other measures, reflecting the importance of perspective in choosing and interpreting efficiency measures. For example, RVUs for a procedure could be used as the input in a measure of the RVUs for physician services delivered during a hospital stay, or as the output of a measure of the RVUs produced by a physician per month.

Figure 2: Outputs Used in Hospital and Physician Efficiency Measures Abstracted from Published Literature

Source: Authors' analysis.

Methodology

Almost half of the measures in the published literature (117 measures, 44.2 percent) were specified as ratios, using single metrics for inputs and outputs. An example of a ratio-based measure is severity-adjusted average length of hospital stay (the ratio of total days of hospital care to discharges, adjusted for patient severity). The rest of the measures in the published literature (147 measures, 55.5 percent) used econometric or mathematical programming methodologies. These approaches differ from ratio-based measures in that they allow for the analysis of multiple metrics of inputs, outputs, and explanatory variables.

Two approaches were most common: data envelopment analysis (DEA) and stochastic frontier analysis (SFA) (Hollingsworth 2003). DEA and SFA belong to a class of methodologies for measuring efficiency called "frontier analysis." Frontier analysis compares a firm's (e.g., hospital, physician practice) use of actual inputs and outputs to efficient combinations of multiple inputs and/or outputs. The two methods use different approaches to calculating the "frontier" of efficient combinations used for comparison. However, DEA, SFA, and ratio-based measures were constructed using similar types of inputs and outputs, typically those in publicly available data. In addition, DEA and SFA require appropriate specification of the relationship between the measured inputs and outputs.

An example of a typical frontier-type measure specification was used by Defelice and Bradford (1997) to compare the efficiency of solo versus group

physician practices. The measure compares the amounts of several inputs—physician labor time, nursing labor time, laboratory tests, and X-rays—used to create a single output, the number of weekly visits per physician. A number of explanatory variables are used including physician-level factors, such as malpractice premiums, years of experience, and nonpractice income; area-level factors, such as the number of hospitals and physicians in the county; and practice-level factors, such as the number of Medicaid patients seen. The measure is calculated using SFA (similar in specification to a linear regression model). Each physician or group practice's efficiency is estimated using a one-sided error term representing the distance from best practice. Because this measure relies on the estimation of a multivariate statistical model, the specification of the model is important. For example, the omission of an important explanatory variable, such as malpractice premiums, could bias the results of the efficiency measurement.

Scientific Soundness

Evidence from tests of reliability and/or validity was reported for six measures (2.3 percent) (Sherman 1984; Vitaliano and Toren 1996; Folland and Hofler 2001; Thomas, Grazier, and Ward 2004). Sensitivity of the measures to elements of specification was reported for 67 measures (25.3 percent). Sensitivity analyses were most common for the measures using multivariate statistical models.

Measures in the Gray Literature

Our scan identified eight major developers of proprietary software packages for measuring efficiency, with other vendors (not shown) providing additional analytic tools, solution packages, applications, and consulting services that build on top of these existing platforms (Table 2). Interviews suggest that purchasers and payers are using vendor-developed measures of efficiency rather than those from the published literature.[1]

Perspective

The vendor-developed measures can be used in multiple applications, but they are most often used by health plans to evaluate the efficiency of physician practices.

Outputs

The vendor-developed measures use two types of health service outputs: (1) episodes of care or (2) care for a defined population over a period of time. In comparison, the majority of measures of physician efficiency in the published

literature used specific types of health services, such as procedures, visits, or hospital discharges, as outputs (Figure 2).

An episode-based approach groups health care services into discrete episodes of care, which are a series of temporally contiguous health care services related to the treatment of a specific acute illness, a set time period for the management of a chronic disease, or provided in response to a specific request by the patient or other relevant entity. The market leaders among episode-based measures are Episode Treatment Groups (ETGs) and Medical Episode Groups (MEGs), which use algorithms primarily based on diagnosis codes and dates of services to group-related insurance claims into episodes (Table 2).

A population-based approach to efficiency measurement classifies a patient population according to morbidity burden in a given period (e.g., 1 year). The outputs, either episodes or risk-adjusted populations, are typically constructed from encounter and/or insurance claims data that include diagnosis and procedure codes, and sometimes pharmacy codes. The measures typically do not directly incorporate the quality of the output.

Inputs

With both episode- and population-based measures developed by vendors, the focus of measure development has mainly been on defining the output of the efficiency measures. Inputs are more loosely defined; typically, inputs are constructed by adding the costs of resources used in the production of that output using administrative data. Cost-based inputs can be constructed using either standardized pricing (e.g., using Medicare fee schedules rather than actual prices paid) or allowing the user to apply actual prices.

Methodology

These measures generally take the form of a ratio, such as observed-to-expected ratios of costs per episode of care, adjusting for patient risk. None of these measures use SFA, DEA, or other regression-based measurement approaches common in the efficiency measures abstracted from the published literature. Almost all of these measures rely on insurance claims data.

Reporting of Scientific Soundness

The grouping algorithms that are the basis of these measures have been used in other applications. For example, most of the population-based measures have been used as risk adjusters for resource utilization prediction, provider profiling, and outcomes assessment. Efforts to validate and test the reliability

Table 2: Efficiency Measures Developed by Private Organizations

Organization	Efficiency Measure Name	Approach	Description
IHCIS-Symmetry of Ingenix	Episode Treatment Groups (ETG)	Episode based	The ETG methodology identifies and classifies episodes of care, defined as unique occurrences of clinical conditions for individuals and the services involved in diagnosing, managing, and treating that condition. Based on inpatient and ambulatory care, including pharmaceutical services, the ETG classification system groups diagnosis, procedure, and pharmacy (NDC) codes into 574 clinically homogenous groups, which can serve as analytic units for assessing and benchmarking health care utilization, demand, and management.
Thomson Medstat	Medical Episode Groups (MEG)	Episode based	MEG is an episode-of-care-based measurement tool predicated on clinical definition of illness severity. Disease stage is driven by the natural history and progression of the disease and not by the treatments involved. Based on the disease staging patient classification system, inpatient, outpatient, and pharmaceutical claims are clustered into approximately 550 clinically homogenous disease categories. Clustering logic (i.e., construction of the episode) includes: (1) starting points; (2) episode duration; (3) multiple diagnosis codes; (4) lookback mechanism; (5) inclusion of nonspecific coding; and (6) drug claims.
Cave Consulting Group	Cave Grouper	Episode based	The CCGroup Marketbasket System compares physician efficiency and effectiveness to a specialty-specific peer group using a standardized set of prevalent medical condition episodes with the intent of minimizing the influence of patient case mix (or health status) differences and methodology statistical errors. The Cave Grouper groups over 14,000 unique ICD.9 diagnosis codes into 526 meaningful medical conditions. The CCGroup EfficiencyCare Module takes the output from the Cave Grouper and develops specialty-specific physician efficiency scores that compare individual physician efficiency (or physician group efficiency) against the efficiency of a peer group of interest.
National Committee for Quality Assurance (NCQA)	Relative Resource Use (RRU)	Population based	The RRU measures report the average relative resource use for health plan members with a particular condition compared with their risk-adjusted peers. Standardized prices are used to focus on the quantities of resources used. Quality measures for the same conditions are reported concurrently.

The Johns Hopkins University	Adjusted Clinical Groups (ACG)	Population based	ACGs are clinically homogeneous health status categories defined by age, gender, and morbidity (e.g., reflected by diagnostic codes). Based on the patterns of a patient's comorbidities over a period of time (e.g., 1 year), the ACG algorithm assigns the individual into one of 93 mutually exclusive ACG categories for that span of time. Clustering is based on: (1) duration of the condition; (2) severity of the condition; (3) diagnostic certainty; (4) etiology of the condition; and (5) specialty care involvement.
3M Health Information Systems	Clinical Risk Grouping (CRG)	Population based	The CRG methodology generates hierarchical, mutually exclusive risk groups using administrative claims data, diagnosis codes, and procedure codes. At the foundation of this classification system are 269 base CRGs, which can be further categorized according to levels of illness severity. Clustering logic is based on the nature and extent of an individual's underlying chronic illness and combination of chronic conditions involving multiple organ systems further refined by specification of severity of illness within each category.
DxCG	Diagnostic Cost Groups (DCG) and RxGroups	Population based	DxCG models predict cost and other health outcomes from age, sex, and administrative data: either or both Diagnostic Cost Groups (DCG) for diagnoses and RxGroups for pharmacy. Both kinds of models create coherent clinical groupings and employ hierarchies and interactions to create a summary measure, the "relative risk score," for each person to quantify financial and medical implications of their total illness burden. At the highest level of the classification system are 30 aggregated condition categories (ACCs), which are subclassified into 118 condition categories (CCs) organized by organ system or disease group.
Health Dialog	Provider Performance Measurement System	Population based	The Provider Performance Measurement System examines the systematic effects of health services resources a person at a given level of comorbidity uses over a predetermined period of time (usually 1 year). The measures incorporate both facility/setting (e.g., use of ER and inpatient services) and types of professional services provided (e.g., physician services, imaging studies, laboratory services). Based on John Wennberg's work, PPMS assesses and attributes unwarranted variations in the system with respect to three dimensions: (1) effective care; (2) preference sensitive care; and (3) supply sensitive care.

Source. Interviews with representatives of organizations.

of these algorithms as tools to create relevant clinical groupings for comparison are typically documented. However, there is very little information available on efforts to validate and test the reliability of these algorithms specifically as efficiency measures. Several studies have examined some of the measurement properties of these vendor-developed measures, but the amount of evidence available is still limited at this time. Thomas, Grazier, and Ward (2004) tested the consistency of six groupers (some episode-based and some population-based) for measuring the efficiency of primary care physicians. They found "moderate to high" agreement between physician efficiency rankings using the various measures (weighted $\kappa = 0.51$–0.73). Thomas and Ward (2006) tested the sensitivity of measures of specialist physician efficiency to episode attribution methodology and cost outlier methodology. Thomas (2006) also tested the effect of risk adjustment on an ETG-based efficiency measure. He found that episode risk scores were generally unrelated to costs and concluded that risk adjustment of ETG-based efficiency measures may be unnecessary. MedPAC has tested the feasibility of using episode-based efficiency measures in the Medicare program (MedPAC, 2006). They tested MEG- and ETG-based measures using 100 percent Medicare claims files for six geographic areas. They found that most Medicare claims could be assigned to episodes, most episodes can be assigned to physicians, and outlier physicians can be identified, although each of these processes is sensitive to the criteria used. The percentage of claims that can be assigned to episodes and the percentage of episodes that can be assigned to physicians were consistent between the two measures. MedPAC also compared episode-based measures and population-based measures for area-level analyses and found that they can produce different results. For example, Miami was found to have lower average per-episode costs for coronary artery disease episodes than Minneapolis but higher average per-capita costs due to higher episode volume.

DISCUSSION

Our systematic review of health care efficiency measures identified a large number of existing measures, the majority appearing in the published literature. The measures in the published literature were typically developed by academic researchers, whereas the measures in the gray literature were developed by vendors and are proprietary. There is almost no overlap between the measures found in the published literature and those in the gray literature, suggesting that the driving forces behind research and practice result in very

different choices of measure. Many of the measures in the published literature rely on methods such as SFA and DEA. The complexity of these methods may have inhibited the use of these measures beyond research, because measurement results can be sensitive to a multitude of specification choices and difficult to interpret. The vendor-developed measures, although relatively few in number, are used much more widely by providers, purchasers, and payers than the measures in the published literature. We observed very little convergence around a "consensus" set of efficiency measures that have been used repeatedly. However, we did observe some convergence around several general approaches—SFA and DEA are most commonly used for research, while ETGs and MEGs appear to have the greatest market share in "real-world" applications. In this way, efficiency measurement differs from quality measurement, where in recent years a fair degree of consensus has been formed around core sets of measures.

The state of the art in health care efficiency measurement contrasts sharply with the measurement of health care quality in several other ways. Unlike the evolution of most quality measures, the efficiency measures in use are not typically derived from practice standards from the research literature, professional medical associations, or expert panels (Schuster, McGlynn, and Brook 1998). Unlike most quality measures, efficiency measures in both the published and gray literature have been subjected to few rigorous evaluations of their performance characteristics including reliability, validity, and sensitivity to methods used. Existing criteria in use could be applied to efficiency measures (e.g., National Quality Forum criteria for quality measures). Measurement scientists would prefer that steps be taken to improve these metrics in the laboratory before implementing them in operational uses. Purchasers and health plans are willing to use measures without such testing under the belief that the measures will improve with use.

A central issue in efficiency measurement is the potential for differences in the quality of the outputs used. Only one of the measures reviewed explicitly incorporated quality of care into the efficiency measure, although several included quality as an explanatory factor in analyses of variation in efficiency or in a separate comparison. Therefore, almost all of the purported efficiency measures reviewed would be classified as "cost of care" measures by the AQA definition, not true "efficiency measures." As discussed above, methods for accounting for quality are not well developed at this time, and preliminary testing shows that results are sensitive to the approach used. These measures could be used under the assumption that variations in quality of care across groups were modest. This assumption may be reasonable for certain

comparisons, for example, several HEDIS measures of cardiac care now show high mean performance with minimal variation. Similarly, the most recent New York State Cardiac Surgery Reporting System results show 38 of 39 hospitals to have between them no statistically significant differences in risk-adjusted mortality ratio. For other types of comparisons, it is likely that quality does vary substantially. Evidence is lacking on the variation in quality for important types of measures, such as cost-per-episode measures that are widely used. If systematic differences in quality do exist, measures that do not account for the differences would reflect the cost of care only, not efficiency. This could provide an incentive for physicians to selectively treat lower-risk patients, potentially increasing disparities in care (Newhouse 1994).

There are several additional unresolved issues in the specification of efficiency measures. These issues include risk adjustment of episode-based measures; attribution of episodes, patients, or other outputs to specific providers when care is dispersed over multiple providers; levels of reliability for detecting differences among entities; and differences between proprietary grouper methodologies (Milstein and Lee 2007). One concern of providers who are the subject of efficiency measurement is that proprietary measures are not sufficiently transparent to allow for definitive answers to these and related measurement questions (American Medical Association 2007). These concerns, along with the quality measurement issue, have been the basis of legal action against users of efficiency measures (Massachusetts Medical Society 2008).

Our findings are subject to several important limitations. We excluded studies from non-U.S. data sources, primarily because we judged the studies done on U.S. data would be most relevant to the task we were contracted to perform. It is possible, however, that adding the non-U.S. literature would have identified additional measures of potential interest. Other exclusion criteria we used may have omitted some relevant studies from the results. For example, our review did not find any measures at the health system level, but several authors have examined "waste" in the U.S. health care system and other issues related to efficiency, estimating that approximately 30 percent of spending is wasteful (Reid et al. 2005).

An important limitation common to systematic reviews is the quality of the original studies. A substantial amount of work has been done to identify criteria in the design and execution of the studies of the effectiveness of health care interventions, and these criteria are routinely used in systematic reviews of interventions. However, we are unaware of any such agreed-upon criteria that assess the design or execution of a study of a health care efficiency

measure. We did evaluate whether studies assessed the scientific soundness of their measures (and found this mostly lacking).

Notwithstanding these limitations, our systematic review suggests that the state of efficiency measurement lags far behind quality measurement in health care. Stakeholders are under pressure to make use of existing efficiency measures and "learn on the job." However, use of these measures without greater understanding of their measurement properties is likely to engender resistance from providers subject to measurement and could lead to unintended consequences. Using measures that do not account for differences in the quality of the outputs being compared could lead purchasers to favor providers who achieve favorable cost profiles by withholding necessary care. Using measures that are not reliable enough to detect differences could lead patients to change physicians, disrupting continuity, without achieving better value for health care spending. Going forward, it will be essential to find a balance between the desire to use tools that facilitate improving the efficiency of care delivery and producing inaccurate or misleading information from tools that are not ready for prime time.

ACKNOWLEDGMENTS

NOTE

1. This conclusion was supported by our interviews with selected stakeholders as well as a study by the Center for Studying Health Systems Change (see Draper, Liebhaber, and Ginsburg 2007).

REFERENCES

Albritton, T. A., M. D. Miller, M. H. Johnson, and D. W. Rahn. 1997. "Using Relative Value Units to Measure Faculty Clinical Productivity." *Journal of General Internal Medicine* 12 (11): 715–7.

American Medical Association. 2007. "Tiered and Narrow Physician Networks" [accessed on June 15, 2007]. Available at http://www.wsma.org/memresources/memresources_pdf/P4P_AMA_Tiered%20_Narrow_Networks_0606.pdf

AQA. 2007. "AQA Principles of 'Efficiency' Measures" [accessed on February 10, 2007]. Available at http://www.aqaalliance.org/files/PrinciplesofEfficiencyMeasurementApril2006.doc

Ashby, J., S. Guterman, and T. Greene. 2000. "An Analysis of Hospital Productivity and Product Change." *Health Affairs* 19 (5): 197–205.

Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Insurance Trust Funds. 2007. "2007 Annual Report" [accessed on June 15, 2007]. Available at http://www.cms.hhs.gov/ReportsTrustFunds/

Bush, R. W. 2007. "Reducing Waste in US Health Care Systems." *Journal of the American Medical Association* 297 (8): 871–4.

Carey, K. 2003. "Hospital Cost Efficiency and System Membership." *Inquiry* 40 (1): 25–38.

Cassel, C. K., and T. E. Brennan. 2007. "Managing Medical Resources: Return to the Commons?" *Journal of the American Medical Association* 297 (22): 2518–21.

Chernew, M., D. M. Cutler, and P. S. Keenan. 2005. "Increasing Health Insurance Costs and the Decline in Insurance Coverage." *Health Services Research* 40 (4): 1021–39.

Conrad, D. A., A. Sales, S. Y. Liang, A. Chaudhuri, C. Maynard, L. Pieper, L. Weinstein, D. Gans, and N. Piland. 2002. "The Impact of Financial Incentives on Physician Productivity in Medical Groups." *Health Services Research* 37 (4): 885–906.

Conrad, D., T. Wickizer, C. Maynard, T. Klastorin, D. Lessler, A. Ross, N. Soderstrom, S. Sullivan, J. Alexander, and K. Travis. 1996. "Managing Care, Incentives, and Information: An Exploratory Look Inside the 'Black Box' of Hospital Efficiency." *Health Services Research* 31 (3): 235–59.

Davis, K., C. Schoen, S. C. Schoenbaum, M. M. Doty, A. L. Holmgren, J. L. Kriss, and K. K. Shea. 2007. "Mirror, Mirror on the Wall: An International Update on the Comparative Performance of American Health Care" [accessed on June 15, 2007]. Available at http://www.commonwealthfund.org/publications/publications_show.htm?doc_id=482678

DeFelice, L. C., and W. D. Bradford. 1997. "Relative Inefficiencies in Production between Solo and Group Practice Physicians." *Health Economics* 6 (5): 455–65.

Dewar, D. M., and J. Lambrinos. 2000. "Does Managed Care More Efficiently Allocate Resources to Older Patients in Critical Care Settings?" *Cost and Quality Quarterly Journal* 6 (2): 18–26.

Draper, D. A., A. Liebhaber, and P. B. Ginsburg. 2007. "High-Performance Health Plan Networks: Early Experiences" [accessed on June 15, 2007]. Available at http://www.hschange.com/CONTENT/929/

Fiscella, K., P. Franks, M. R. Gold, and C. M. Clancy. 2000. "Inequality in Quality: Addressing Socioeconomic, Racial, and Ethnic Disparities in Health Care." *Journal of the American Medical Association* 283 (19): 2579–84.

Fisher, E., D. Wennberg, T. Stukel, D. Gottlieb, F. Lucas, and E. Pinder. 2003a. "The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care." *Annals of Internal Medicine* 138 (4): 273–87.

———. 2003b. "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care." *Annals of Internal Medicine* 138 (4): 288–98.

Folland, S., and R. Hofler. 2001. "How Reliable Are Hospital Efficiency Estimates? Exploiting the Dual to Homothetic Production." *Health Economics* 10 (8): 683–98.

Fuchs, V. 2005. "Health Care Expenditures Reexamined." *Annals of Internal Medicine* 143 (1): 76–8.

Harrison, J. P., and M. N. Coppola. 2007. "The Impact of Quality and Efficiency on Federal Healthcare." *International Journal of Public Policy* 2 (3): 356–71.

Hollingsworth, B. 2003. "Non-Parametric and Parametric Applications Measuring Efficiency in Health Care." *Health Care Management Science* 6 (4): 203–18.

Iglehart, J. K. 2002. "Changing Health Insurance Trends." *New England Journal of Medicine* 347 (12): 956–62.

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

Kessler, D. P., and M. B. McClellan. 2002. "The Effects of Hospital Ownership on Medical Productivity." *RAND Journal of Economics* 33 (3): 488–506.

The Leapfrog Group and Bridges to Excellence. 2007. "Measuring Provider Efficiency Version 1.0" [accessed on January 5, 2009]. Available at http://www.google.com/url?sa=t&source=web&ct=res&cd=4&url=http%3A%2F%2Fbridgestoexcellence.org%2FDocuments%2FMeasuring_Provider_Efficiency_Version1_12-31-20041.pdf&ei=BGdiSduROZWksAOj0NyODQ&usg=AFQjCNHFoeZecG_99bd3fmSFaEkBDAKKNA&sig2=UoU5OCgfaXePw2Yp8h5xeA

Massachusetts Medical Society. 2008. "MMS Files Legal Action Against the Group Insurance Commission, Its Executive Director, and Two Health Plans" [accessed on May 31, 2008]. Available at http://www.massmed.org/AM/Template.cfm?Section=Home&TEMPLATE=/CM/ContentDisplay.cfm&CONTENTID=21943

McGlynn, E., P. Shekelle, S. Chen, G. Goldman, J. Romley, P. S. Hussey, H. de Vries, M. Wang, M. Timmer, J. Carter, C. Tringale, and R. Shanman. 2008. "Health Care Efficiency Measures: Identification, Categorization, and Evaluation." AHRQ Publication No. 08-0030 [accessed September 16, 2008]. Available at http://www.ahrq.gov/qual/efficiency/index.html

Medicare Payment Advisory Commission (MedPAC). 2006. *Report to the Congress: Increasing the Value of MEDICARE.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission (MedPAC). 2007. *Report to Congress: Assessing Alternatives to the Sustainable Growth System.* Washington, DC: MedPAC.

Milstein, A. 2004. "Hot Potato Endgame." *Health Affairs* 23 (6): 32–4.

Milstein, A., and T. H. Lee. 2007. "Comparing Physicians on Efficiency." *New England Journal of Medicine* 357 (26): 2649–52.

Newhouse, J. P. 1994. "Frontier Estimation: How Useful a Tool for Health Economics?" *Journal of Health Economics* 13 (3): 317–22.

O'Kane, M., J. Corrigan, S. M. Foote, S. R. Tunis, G. J. Isham, L. M. Nichols, E. S. Fisher, J. C. Ebeler, J. A. Block, B. E. Bradley, C. K. Cassel, D. L. Ness, and J. Tooker. 2008. "Crossroads in Quality." *Health Affairs* 27 (3): 749–58.

Reid, P. P., W. D. Compton, J. H. Grossman, and G. Fanjiang, eds. 2005. *Building a Better Delivery System: A New Engineering/Health Care Partnership.* Washington, DC: National Academies Press.

Rosenman, R., and D. Friesner. 2004. "Scope and Scale Inefficiencies in Physician Practices." *Health Economics* 13 (11): 1091–116.

Rosko, M. D. 2004. "Performance of US Teaching Hospitals: A Panel Analysis of Cost Inefficiency." *Health Care Management Science* 7 (1): 7–16.

Schuster, M. A., E. A. McGlynn, and R. H. Brook. 1998. "How Good Is the Quality of Health Care in the United States?" *The Milbank Quarterly* 76 (4): 517–63.

Sherman, H. D. 1984. "Hospital Efficiency Measurement and Evaluation. Empirical Test of a New Technique." *Medical Care* 22 (10): 922–38.

Skinner, J., E. Fisher, and J. E. Wennberg. 2007. "The Efficiency of Medicare" [accessed on June 15, 2007]. Available at http://ssrn.com/abstract=277305

Thomas, J. W. 2006. "Should Episode-Based Economic Profiles Be Risk Adjusted to Account for Differences in Patients' Health Risks?" *Health Services Research* 41 (2): 581–98.

Thomas, J. W., K. L. Grazier, and K. Ward. 2004. "Economic Profiling of Primary Care Physicians: Consistency among Risk-Adjusted Measures." *Health Services Research* 39 (4, part 1): 985–1003.

Thomas, J. W., and K. Ward. 2006. "Economic Profiling of Physician Specialists: Use of Outlier Treatment and Episode Attribution Rules." *Inquiry* 43 (3): 271–82.

Timbie, J. W., and S. L. Normand. 2007. "A Comparison of Methods for Combining Quality and Efficiency Performance Measures: Profiling the Value of Hospital Care Following Acute Myocardial Infarction." *Statistics in Medicine* 27 (9): 1351–70.

Vitaliano, D. F., and M. Toren. 1996. "Hospital Cost and Efficiency in a Regime of Stringent Regulation." *Eastern Economic Journal* 22: 161–75.

Weingarten, S. R., L. Lloyd, C. F. Chiou, and G. D. Braunstein. 2002. "Do Subspecialists Working Outside of Their Specialty Provide Less Efficient and Lower-Quality Care to Hospitalized Patients Than Do Primary Care Physicians?" *Archives of Internal Medicine* 162 (5): 527–32.

The White House. 2007. "Executive Order: Promoting Quality and Efficiency Health Care in Federal Government Administered or Sponsored Health Programs" [accessed on July 1, 2007]. Available at http://www.whitehouse.gov/news/releases/2006/08/20060822-2.html

Zuckerman, S., J. Hadley, and L. Iezzoni. 1994. "Measuring Hospital Efficiency with Frontier Cost Functions." *Journal of Health Economics* 13 (3): 255–80; discussion 335–40.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.