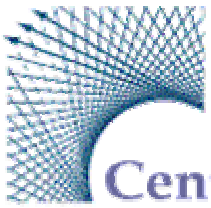


Designing a Pay for Performance Methodology for Maryland Hospitals

Presented to the Maryland Health Services Cost Review Commission

Center for Performance Sciences

December 2004

Table of Contents

SECTION I. EXECUTIVE SUMMARY 1

Introduction 1

About the Authors of this Report 1

Review of Recommendations 1

How Quality Will Be Measured 4

How Quality and Quality Improvement Will Be Encouraged Through Payments 4

Rewards and Incentives 5

Infrastructure Support 6

Determining Level and Source of Funding 7

SECTION II. TECHNICAL REPORT 8

A Look at Other Industries 8

Framework Development 12

Dimensions of Quality 12

Measurement Considerations 15

Measuring Progress 15

Measuring Experience 16

Measuring Productivity and Safety 18

Payment Considerations 21

How to Compare Hospitals 25

Possible Limitations 28

Direct Care Provider Involvement and Support 28

Accommodating Patient and Community Preferences 29

Rapidly Changing Technology 29

The Impact to Patient Care, Hospitals, and the Payment System 30

Conclusion 32

RECOMMENDATIONS 34

Recommendation 1: Decision on Terminology 34

Recommendation 2: The Structure and Composition of the Evaluation Workgroup .. 34

Recommendation 3: Prioritization of Focus Areas 34

Recommendation 4: Implementation Phases 34

Recommendation 5: Beyond Financial Factors 35

Recommendation 6: Quantitative and Qualitative Measurement Strategies 35

Recommendation 7: Hospital Cluster Analysis 36

Recommendation 8: Comparisons Using Baseline References 36

Recommendation 9: Direct Care Providers as a Target Group 36

Recommendation 10: Lessons from Other Industries 36

List of Figures

Figure 1: Pay for Performance Efforts by Type of Sponsor 2

Figure 2: Review of Key Recommendations – Benefits and Potential Risks 3

Figure 3: Lessons from Other Industries for HSCRC Pay-for-Performance 9

Figure 4: Dimensions of Quality and Their Measurement 13

Figure 5: Frequency of Use of Categories of Measures in Pay-for-Performance 14

Figure 6: Major Measurement Considerations 16

Figure 7: Additional Measurement Considerations for Productivity and Safety 18

Figure 8: Phasing In Dimensions Based on Ability to Measure 21

Figure 9: Major Payment Considerations At This Time 21

Figure 10: Scenario 1 – "Ongoing Improvement" 23

Figure 11: Scenario 2 – "Continued Improvement" 24

Figure 12: Scenario 3 – "Breakthrough" 24

Figure 13: A Framework for Comparing Hospitals 27

Figure 14: A Way to Measure Progress 28

Figure 15: Payment System Impact 30

Figure 16: Pay for Performance and Patient Care 31

Figure 17. Prioritization of focus areas and the sequence of their phasing into a pay-for-performance model 34

Figure 18: Future Phases and Activities 35

SECTION I. EXECUTIVE SUMMARY

Introduction

As health care costs continue to grow, payers, purchasers, and providers have begun to explore different ways to manage health care expenditures and allocate available resources. In parallel, the focus on issues such as patient safety and quality of care has also increased as the research matures on these topics. The confluence of these pressures has led to an increased interest in pay-for-performance and incentives. Maryland's exploration of this concept is timely as private organizations, public payors, and others have begun to experiment with the idea.

This paper reviews the recommendations of a Steering Committee convened by the HSCRC in 2003 on the topic of pay-for-performance. These recommendations are examined from the perspective of how to measure quality and how to encourage quality and its improvement through incentive and reward payment algorithms. Furthermore, this paper considers other industries' experiences and lessons learned with pay-for-performance approaches.

In addition to a review of the HSCRC's Steering Committee's recommendations, this paper initiates the process of developing a framework through which to consider pay-for-performance. The framework identifies dimensions of quality and key considerations for measurement and payment. It also creates a construct to view them relative to one another at a hospital level. Finally, this framework examines how pay-for-performance fits within the broader health care system and affects patient care.

About the Authors of this Report

The HSCRC commissioned Vahé A. Kazandjian, PhD, MPH, as the principal author for this report. Dr. Kazandjian is the President of the Center for Performance Sciences (CPS), a subsidiary of the Maryland Hospital Association (MHA), and Professor (Adjunct) at the Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland.

Review of Recommendations

The HSCRC Steering Committee's report of February 2004 provided an overview of the history of quality measurement nationally and in Maryland and the more recent activity around establishing pay-for-performance in health care. Based on this overview and its discussions, it set forth the following recommendations:

1. Commence with a pay-for-performance system in Maryland hospitals "as soon as reasonably possible" as "considerable comfort" exists in assessing quality

2. Adopt mission, vision, and goals that focus on improving quality, effectiveness, and efficiency of health care by providing funding

3. Divide funding/payment to hospitals into three streams: (1) for performance above a set threshold; (2) for performance that is below the set threshold but

meets an improvement target; and (3) for infrastructure support when performance does not meet the set threshold and is driven by a lack of appropriate infrastructure

4. Examine available measures (process, patient and performer safety, outcome, patient satisfaction and experience, etc.) to recommend an initial set of measures

5. Craft a process in the future to evaluate how measures are meeting the pay-for-performance initiative's goals and consider new measures

6. Create a data collection policy to allow accomplishment of the pay-for-performance initiative's goals while limiting additional reporting burden on hospitals

7. Develop a composite scoring system of the accepted measures from which to base funding/payment

8. Recruit and hire a staff person at the HSCRC to manage the initiative

9. Establish two workgroups: one to develop and design the system and another to evaluate it on an on-going basis

10. Ensure that funding and payments are "significant enough to encourage the behavior that will result in quality outcomes" by determining policies that consider cost-benefit, potential return on investment, system performance, etc.

Before discussing these recommendations more in depth, the importance of moving forward with a pay-for-performance initiative in Maryland should be highlighted. The HSCRC's Steering Committee correctly perceived that the momentum behind pay-for-performance is growing. As health care costs continue to rise and consumerism continues to take hold, payors and purchasers continue to pursue ways to get more "bang for their buck" and focus on quality of care. This can be seen in the doubling of the number of sponsors developing a pay-for-performance effort in the last year.

Figure 1: Pay for Performance Efforts by Type of Sponsor

Sponsor Types	2003	2004
Commercial Health Plans	32	56
Medicaid Plans	1	9
CMS Initiatives	1	5
Employers/Employer Coalitions	5	6
Other	0	4
TOTAL	39	80

Source: Med-Vantage, Inc. National P4P Survey. 2004. Presented at Second Annual National Health Information Technology Summit. Washington, DC. October 20, 2004.

Viewed through the lens of designing and operationalizing a pay-for-performance system, five of the Steering Committee's recommendations stand out. They stand out because they drive two important design themes: (1) how quality will be measured and (2) how quality and quality improvement will be encouraged through payments.

1. Examine available measures and recommend an initial set of measures
2. Develop a composite scoring system
3. Craft a process in the future to evaluate how measures are meeting the pay-for-performance initiative's goals and consider new measures
4. Divide funding/payment to hospitals into three streams
5. Ensure that funding and payments are "significant enough to encourage the behavior that will result in quality outcomes"

Figure 2: Review of Key Recommendations – Benefits and Potential Risks

Concept	Benefits	Potential Risks	Some Potential Mitigating Actions
How to Measure Quality	• Areas of focus seem appropriate • Can accommodate new research • Allows flexibility for a nuanced approach • Can correct or make changes mid-course	• Could entail heavy effort to administer (hospitals and HSCRC) • Potential for too much or inappropriate focus on specific area(s)	• Consider data burden • Further define evaluation work group (EWG) structure and scope • Appropriate mix of skills and expertise on EWG • Active HSCRC staff management
How to Encourage Quality and Quality Improvement Through Payment	• Dynamic • Focuses on sustainable change • Rewards and incentives look at "what," are measurable, and have been shown to work	• Rewards, incentives, and infrastructure payments need to be combined for optimum effect (i.e., continuous quality improvement) • "Why" data not as easy to collect as	• Options for methodology to evaluate "why" need to be developed and tested in conjunction with methodology for "what" • Trend data to be analyzed before

Concept	Benefits	Potential Risks	Some Potential Mitigating Actions
	• "Infrastructure" looks at "why" performance can be achieved and can remove barriers to performance	"what" data	implementation

How Quality Will Be Measured

As will be seen in the discussion on framework development, how quality will be measured is a somewhat circular theme. With the relative amorphousness involved in defining quality currently, how quality is measured often determines what quality is. The approach set forth by the February HSCRC Steering Committee recommendations accommodates the evolutionary nature of the thinking on health care quality and its measurement. In other words, by developing an initial set of measures and by creating a process to evaluate these on a periodic basis, new research and thinking can be integrated into the system. Furthermore, continuing re-evaluation allows the system to shift its focus on different dimensions or aspects of those dimensions as appropriate. This provides the pay-for-performance system with the ability to target specific areas of improvement, hone in on different nuances or flavors of those areas of improvement, and make course corrections mid-stream if needed.

The primary potential concern of taking an iterative and evolving approach is the administrative effort needed to select, develop, track, and maintain the appropriate metrics and scoring systems. This approach does not lend itself to "auto-pilot" and will require active monitoring and management on the part of HSCRC staff. Furthermore, the data burden on hospitals would need to be considered as the system and what is measured evolves. More deliberation will need to occur on the structure and composition of the evaluation group, the frequency with which it will be re-constituted, and the scope of its work to ensure that it can provide the appropriate guidance and support for the evolution of the system.

Another possible drawback is that the system could "over-focus" or inappropriately focus on one dimension of quality or one aspect of one dimension of quality. This concern was expressed by the Steering Committee in their desire to create an ongoing process of evaluation. Diligence and an appropriate mix of skills and expertise on the part of the evaluation work group will go a long way to mitigate this concern.

How Quality and Quality Improvement Will Be Encouraged Through Payments

How quality and quality improvement will be encouraged through payments is the second major theme of design captured by the HSCRC Steering Committee's recommendations.

Here too the dynamic nature of the proposed approach to measurement will be beneficial. Mid-course assessments and corrections will be crucial. A pilot test, suggested by the HSCRC, would also be useful in determining if the payment structure is appropriate for what is measured and what is measured is appropriate to the goals of the initiative.

The recommendations with regard to payment and funding highlight an important point: sustainability of desired change. Aligning the level of funding with the identified need is necessary so that the monies provided would either generate sustainable change or create an environment in which sustainable change could take hold. Clearly, the HSCRC will need to consider a number of different issues to determine what policies will need to be established to do this. However, it is an important principle to build into the foundation of the system so that it achieves a meaningful and lasting impact.

The three streams of funding recommended – one for achieving a threshold, another for demonstrating improvement, and another for infrastructure support when it becomes evident that a threshold or improvement cannot be attained due to infrastructure issues – cast a broad net. All three are meant to encourage and enable hospitals to change their behavior and improve quality.

Rewards and Incentives

The first two are types of funding more traditionally seen in pay-for-performance systems that have been demonstrated to promote a certain type of behavior. Hospitals have either already achieved that behavior through making a change or they understand that they are expected to change that behavior to an expected level. The "carrots" of payment create rewards and incentives for both. The beauty of a pay-for-performance system that re-evaluates what it measures periodically is that a "stick" exists for maintaining the status quo. Once hospitals have achieved all performance thresholds, new rewards and incentives are created by setting new thresholds or identifying new areas to measure.

Advantages and disadvantages exist with these two types of funding. Typically, these two types of funding are quantitatively-oriented – an advantage when it comes to the process of collecting and analyzing data. Metrics, thresholds, and improvement levels are numerically set, calculated, and evaluated. Data collection and analysis once these are determined can be more streamlined. Another key advantage is that this approach has been shown to work in health care. The Premier Hospital Quality Incentive Demonstration, part of the CMS Hospital Quality Initiative launched in 2003, is a three-year project that will provide financial rewards across the three years to hospitals in the top twenty percent of quality for clinical areas being measured and will adjust payments for hospitals who fail to improve from their demonstration baseline in year three.

However, neither of these advantages is straightforward. There is some debate as to whether or not providing payments for improvement promotes the acceptance of less-than-optimal quality of care. The counter-argument suggests that solely providing payments for performance above thresholds does not establish a mindset of continuous quality improvement and could promote stagnation and also less-than-optimal quality care. Most importantly, however, the level of detail in the data may not capture why an organization is unable to achieve performance thresholds.

Infrastructure Support

The third type of payment – for infrastructure support when less than optimal level of performance is demonstrated and correlated with a lack of infrastructure – delves deeper than the other two types of payments and begins to address some of the disadvantages they pose. This concept – that to pay for performance, one must first pay for the people, processes, and technologies required for a health care provider to achieve performance – is gaining concurrence in those working on developing pay-for-performance programs.

In health care, unlike in some other industries, one reason this is so important is that many of the required infrastructural changes are of a public good nature. In other words, the majority of the benefits do not necessarily accrue to those who invest the time and resources in making the improvements. This is particularly true in the area of health information technologies such as electronic health records and computerized provider order entry. In fact, a study by the Center for Information Technology Leadership in Massachusetts indicates that over 70 percent of the net benefits for healthcare information exchange and interoperability (HIEI) once it is fully in place would accrue to stakeholders other than providers.[1]

In the area of information technology and information exchange, no one approach to this has been settled on, but payors, purchasers, and providers have been exploring a variety of different approaches help providers bring on those infrastructure improvements that would improve quality but are of a public good nature. For instance, Bridges to Excellence is a set of programs started by a purchaser coalition that provides physicians and patients with incentives to adopt systematic processes to improve quality of care. Part of this effort is to provide physicians with an incentive to invest in the necessary infrastructure and workflow improvements. Another example is the effort by WellPoint to provide its networked physicians with access to free hardware and software that enables the exchange of patient information electronically. This has met limited success as, according to Leonard Schaeffer, WellPoint Health Networks CEO, in the words of former IOM President Harvey Fineberg, MD, PhD, "Free isn't cheap enough."[2] In Massachusetts, Blue Cross Blue Shield of Massachusetts is providing the Massachusetts e-Health Collaborative with up to $50 million to buy and install interoperable electronic medical records software with clinical decision support in physician offices in three communities.[3]

Whereas the other two payment streams look at "what," (i.e., what performance is or what the improvement is), this payment stream looks at "why." It asks: why is a hospital unable to reach a performance threshold? By doing so, it forces an examination of the structures in place at a hospital that may prevent it from achieving improvement or

---

[1] Middleton, B. Center for Information Technology Leadership, Boston, MA. Assessing Value/Calculating ROI. Presentation at the Second Annual National Health Information Summit. Washington DC, October 20, 2004.

[2] Schaeffer, L. WellPoint Health Networks, Thousand Oaks, CA. Transforming an IT-Enabled Health Care System: The Health Plan Role. Presentation at the Second Annual National Health Information Summit. Washington DC, October 20, 2004

[3] Chin, T. Proposed Massachusetts e-Health Network Gets $50 Million Boost. Amednews.com. American Medical Association. October 18, 2004.

targets. These structures may be information technology but can go beyond information technology. They may be the structures that look at how information and knowledge flow across actors in the system or they may be what are thought of traditionally as infrastructure, like physical plant and equipment. In other words, these payments will allow hospitals to assure that their people, processes, and technologies are at a point where they can participate effectively in a reward and incentive program that seeks to promote continuous quality improvement.

Again, this type of funding brings its own benefits and pitfalls. If measurement is done appropriately, this mechanism can concentrate funding on the barriers to performance at a given hospital. In fact, one could argue that this is more effective theoretically than simply providing payments for meeting performance thresholds or demonstrating improvement. It is a basic tenet of management that when performance is not as anticipated or expected barriers to performance should be identified and removed. However, the challenge in a pay-for-performance system in health care is that measuring "why" performance is not as expected is not straightforward. At a high level, the measurement of "why" is less of a quantitative exercise than one of qualitative exploration. As a result, data challenges will naturally arise. From a payment standpoint, this presents difficulty because why performance does not meet thresholds will vary from hospital to hospital. Additionally, funding one aspect of infrastructure correlated with enhance performance at one hospital may have a different result at another hospital. Only by collecting enough information to examine trends will the HSCRC be able to have confidence in what it is measuring and setting payment levels for.

Determining Level and Source of Funding

Clearly, the HSCRC will need to identify a source for these three streams of funding. The parameters set forward by the HSCRC's Steering Committee as to what to consider in determining the level and source of funding are a good start. These included system performance (e.g., Medicare waiver test and all-payor test), cost-benefit analysis (i.e., whether and how savings will accrue at a system level), and return on investment. In terms of return on investment, however, it is important to note that the concept needs to be expanded beyond traditional financial terms. ROI in the quality field may very well need to be calculated in terms of improved compliance with health maintenance protocols, improvement adherence to clinical guidelines, reductions in adverse events and near misses, etc. The HSCRC could also consider comparing Maryland hospitals to the nation with regard to their performance on quality indicators. This becomes more possible as data collection efforts nationwide are developed and refined.

SECTION II. TECHNICAL REPORT

This section presents the systematic review of the HSCRC recommendations and provides a rationale for the prioritization of focus areas and next phases discussed in the Executive Summary.

While this paper has based its recommendations on available experiences from the health care and other industries, the authors synthesized these experiences to address issues specific to the health care environment in Maryland.

A Look at Other Industries

In health care, pay for performance is often equated for pay for quality. These concepts are congruous but not quite identical. In fact, many pay for performance programs explicitly are about improving quality and economic performance. For example, requiring physicians to prescribe from a formulary has implications both for quality and financial performance. The savings generated from this can go towards subsidizing incentives in other areas.

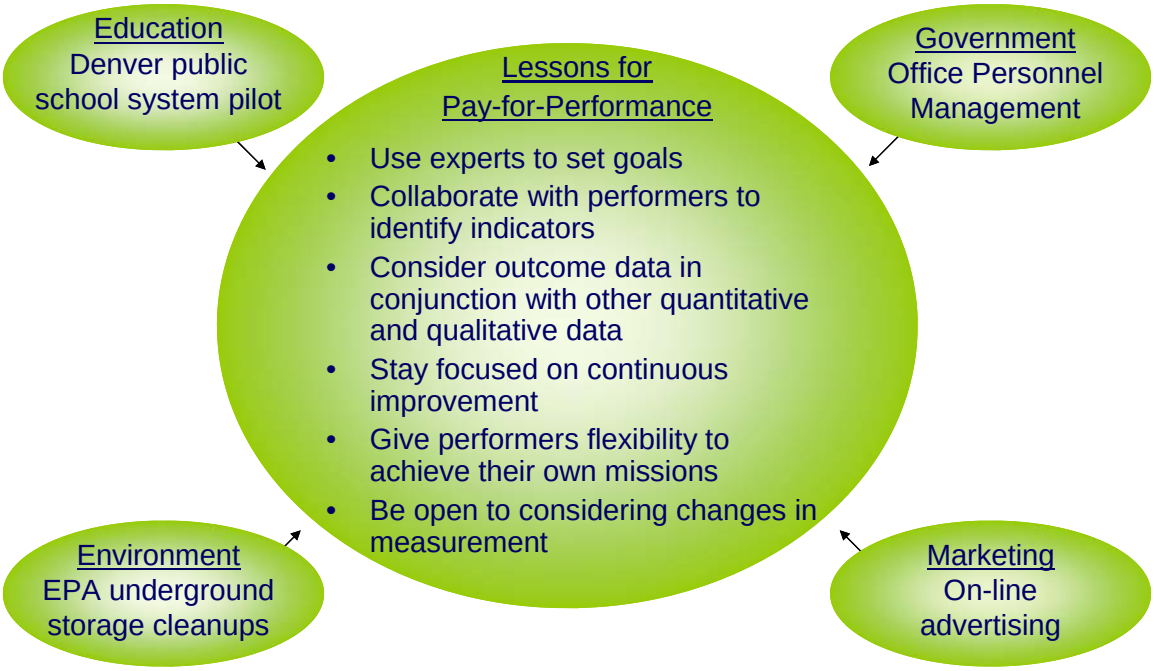
The parallel themes of quality and financial performance run throughout other industries' efforts to pay for performance. For instance, pay-for-performance is a common practice in the private sector. Compensation experts believe that roughly two-thirds of US companies use some sort of variable pay techniques, affecting about ten percent of overall compensation. Recent trends indicate that companies are becoming more selective about who gets incentive pay as companies shift their focus from rewarding past performance to motivating continued improvement in future performance.[4]

Furthermore, a basic concept of contracting is the performance guarantee. This takes the concept of paying for services one step further, creating incentives for the contracted vendor to meet and surpass performance goals.

Other examples can be drawn from education, the government, the environment, and marketing for lessons for how to design a successful pay-for-performance program.

[4] Bates, S. (Top) Pay for (Best) Performance. HR Magazine. 2003 Jan: 48(1).

Figure 3: Lessons from Other Industries for HSCRC Pay-for-Performance



The most common example cited in pay-for-performance is in education. Merit pay for teachers has been used throughout the twentieth century, and paying teachers based on how well they perform is gaining increased support as a way to improve the quality of education and teaching throughout the country. State legislatures have seen it as a key policy tool; 30 of 50 states have passed legislation requiring some type of performance-based pay for teachers.[5] Advocates of pay-for-performance for teachers view the uniformity of teachers' salaries as a barrier to attracting and retaining talent. Interestingly, this leveling of teacher salaries was originally introduced to reduce the inequities in pay between female and male teachers, white and minority teachers, elementary and high school teachers, and so forth. Opponents of merit pay fear that what is measured may not be what is important and that pay-for-performance may undermine the collaborative nature of the profession.

Despite the increasing support, pay-for-performance programs in education have met with mixed success and much controversy. Some key features of successful merit pay programs includes rewarding both group performance with respect to a group of students as well as individual teachers, encouraging individuals to expand professional responsibilities (e.g., "career ladders"), creating non-threatening opportunities for feedback and remediation, and using it as a proportion of total compensation. Certain merit pay programs define performance not just as teachers' activities, but student achievement. For example, the Denver public school system has a fifteen school pilot program that is testing three approaches. The first looks at standard achievement tests.

[5] Delisio, ER. Pay for Performance: What are the Issues? Education World. Jan 27, 2003.

The second allows teachers, working with their school principals, to define two objectives based on the academic achievement of their students. The third looks at individual teachers' portfolios that include student achievement scores.[6]

Education is not the only public arena that is experimenting with and using merit pay. The federal government has also struggled to link performance with compensation. The chronology is complex. The idea of pay-for-performance was introduced as early as 1883 with the Civil Service Act which required promotions by merit, but did not establish an overall appraisal system. It was not until 1935 that one was initially established, and it continued to be modified until 1978, when the Civil Service Reform Act required that agencies develop appraisal systems for all employees and establish performance-related pay under a Merit Pay System. This was eventually replaced by the Performance Management and Recognition System (PMRS). PMRS was standardized throughout the 1980s, and eventually terminated in 1993 due to dissatisfaction with its "one-size-fits-all" approach. Since then, performance management has been decentralized to the agency level so that they may best develop programs to meet their specific needs. Agencies are required to develop strategic plans and performance plans for program activities that establish objective, quantifiable, and measurable goals, determine performance indicators, and compare program results with plan goals. The focus now is on the agency and program's success and how individual employees and groups of employees contribute to that success.[7]

Merit pay is not the only manifestation of pay-for-performance. The US Environmental Protection Agency (EPA) has introduced pay-for-performance as an alternative approach to the traditional "time and materials" approach to contracting for cleanups of leaking underground storage tank (LUST) sites. The EPA did not replace the traditional approach, but encouraged states to look at their sites and determine where pay-for-performance contracting would result in faster and better cleanup efforts. As of September 2001, the EPA had confirmation of roughly 425,000 LUST sites, sixty percent of which had cleanups completed. For the over 170,000 cleanups remaining, the EPA's goal was to reduce the number of cleanups as expeditiously as possible and potentially lowering the cost of the cleanup efforts without compromising their quality. Rather than having contractors bill for the hours of work done and cost of materials needed to complete a cleanup, the pay-for-performance approach pays contractors a set amount for reaching predetermined contamination reduction goals within a given time limit. Escape clauses are also included in this approach to provide contractors with recourse so that they may reduce the risk they take on. This created an incentive for contractors to meet clean up goals as quickly as possible as well as contain the costs associated with materials and other expenditures. Early reports from the EPA indicated that this approach was leading to use of more innovative and aggressive approaches and technologies as well as reducing the administrative burden of oversight (e.g., monthly invoice comparison). As of July 2003, fifteen states across the country had already adopted pay-for-performance in some or all of their LUST cleanups. Twenty-two other

[6] Education Commission of the States. Issue Paper – Pay-for-Performance: Key Questions and Lessons from Five Current Models. Jun 2001.
[7] US Office of Personnel Management. Performance Management Guidance. Sep 2001.

states were considering it, with seven of these twenty-two (including Maryland) ready to conduct pilot projects using pay-for-performance.[8]

Finally, marketing has also embraced pay-for-performance. As online advertising has become more popular, advertisers have begun to measure the effectiveness (quality) of their campaigns at different levels. A commonly used metric to date has been cost per thousand impressions, which looks at the number of times a particular ad has been viewed. More recently, advertisers have begun to look at whether viewers of their ads are taking specific actions (e.g., a sales transaction as a result of the ad). This has led to the development of cost per action as a common basis for advertising campaigns. Although cost per action has been long used as a way to measure the effectiveness (quality) of direct response mailing advertisement campaigns, its adoption rate in the online world has been slower as the sophistication of consumers and advertisers has had to evolve in order to make the metric and its interpretation accurate. Analysts of the industry predict that roughly one-third of on-line advertising revenue will be paid on a pay-for-performance basis by 2006.[9]

In each of these examples, performance is defined by the person/people performing the work and by the consequent results and outcomes. These systems look at whether the performers are doing the "right thing" and how well they do it by looking at the efficiency with which they do the "right thing" and the impact of their actions on the people and situations meant to be affected. In doing so, measurement is clearly one of the most contentious issues. Experts should be brought in to set goals where possible, as with the EPA's use of state experts in determining specified decontamination levels for their LUST cleanup sites. Collaboration between performers and those who manage or regulate them can also be a critical element in determining what to measure, as shown by the Denver public school system's pilot. Data on outcomes and results (e.g., student achievement scores) should be considered in conjunction with data on what the performer is doing and how efficiently it is being done (e.g., teacher portfolio). Focus should be kept on continuing improvement and progress but organizations should have flexibility to work towards their own strategic goals, a lesson learned by the federal government in its struggle to reform its own performance measurement system. New ways of doing things, like advertising on-line, do not necessarily need new ways of measurement, but the science of measurement may need time to evolve so that it can be accurate to be believed.

Using experts to set goals, collaborating with performers to identify indicators, considering outcome data with other quantitative and qualitative data, staying focused on continuous improvement, giving performers flexibility to achieve their own missions, and being open to considering changes in measurement – these are important lessons to keep in mind as Maryland moves forward with developing its pay-for-performance system for its hospitals.

[8] US Environmental Protection Agency. Pay for Performance Toolbox. www.epa.gov/swerust1/pfp/toolbox1.htm

[9] Riolo, L. Arguing for Pay for Performance. Media Life. Sep 3, 2002.

Framework Development

In reviewing the five key design recommendations to determine how best to operationalize them, it is important to place them in the context of how quality health care is defined. As the HSCRC Steering Committee report indicates, quality health care is a difficult and somewhat amorphous concept to define. AHRQ's definition of the right thing at the right time in the right way for the right person with the best results possible is juxtaposed with the IOM definition of "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge." This is why identifying and defining the dimensions of quality that will be addressed is important.

Any dimensions, however, will have issues with measurement. Well-defined measurement, both of the progress that can be achieved and the actual experience of Maryland hospitals with the dimensions of quality, is critical. We will explore some of the issues and examine what types of measures may be better suited for the different dimensions, specifically focusing on productivity and safety. We will also lay out some key considerations in measuring actual experience that can be used as criteria in the measurement selection process. These considerations provide information the state of the art with regard to measuring the different dimensions and drive the order that the HSCRC should consider the different dimensions. In addition, we will also outline some of the key options for projecting expected performance against which to compare actual experience to determine the amount of progress being made.

Having identified the dimensions of quality to be addressed, it also becomes possible to think about how they relate to one another at the hospital level. The ability to compare hospitals at a point in time is necessary to be able to identify which hospitals qualify for which type of payment and how much. Over time, ideally, the HSCRC will be able to use the data captured in this comparison model to be able to look at the amount of change achieved as a result of the overall system as well as for the individual funding streams.

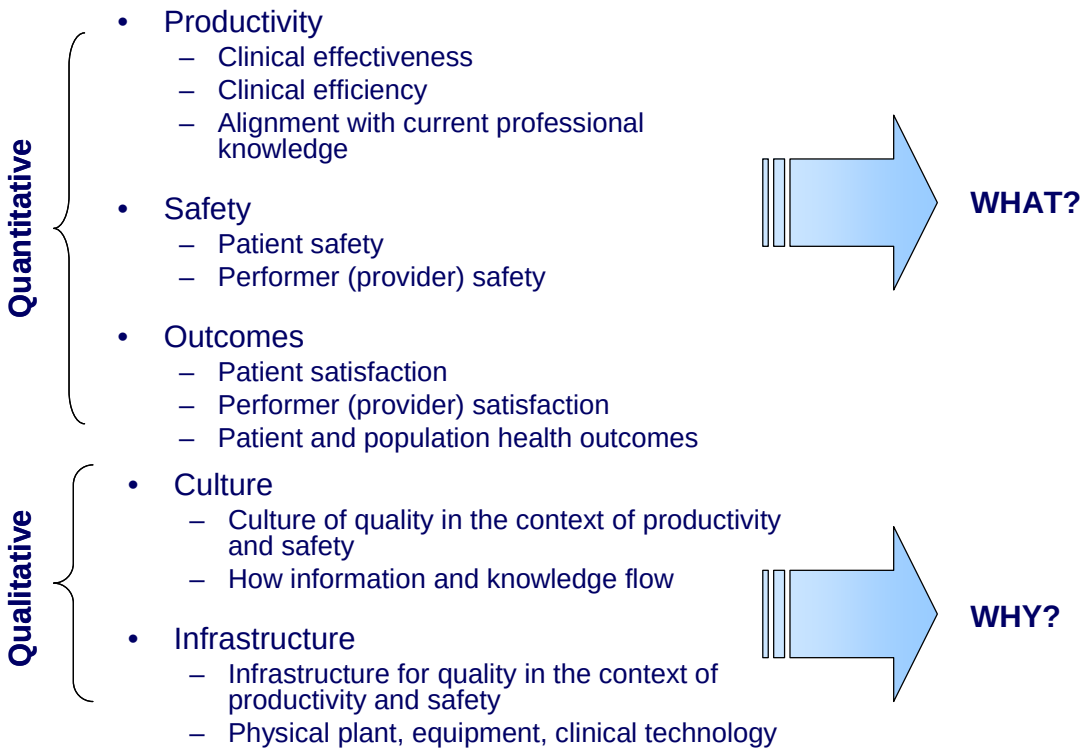
Dimensions of Quality

By breaking apart the payment streams into rewards, incentives, and infrastructure support, there are two primary things that the Steering Committee implicitly recommends to be measured and changed with regard to quality: what is happening in hospitals and why.

Several classification schemes exist to think about indicators, the most well-known being structure, process, and outcomes.[10] Looking at this triumvirate in the context of "what" and "why" produces multiple dimensions of quality that the HSCRC pay-for-performance initiative can influence. In fact, doing so allows for important concepts such as safety and culture to be studied independently rather than as part of structure, part of process, and part of outcomes.

[10] Donabedian A. Explorations in Quality Assessment and Monitoring Volume I: The Definition of Quality and Approaches to Its Assessment. Health Administration Press. 1980.

Figure 4: Dimensions of Quality and Their Measurement



To understand "what" quality exists, productivity, safety, and outcomes can be analyzed. Future work will identify specific measures in each of these dimensions. Broadly speaking, however, productivity refers to clinical effectiveness, clinical efficiency, and alignment of care with current professional knowledge (e.g., care protocols, care guidelines, evidence-based medicine, etc.). Safety should be considered both from the perspective of the patient as well as the caregiver. The concept of outcomes includes both traditional patient and population health outcomes and satisfaction of both the patient and performer (provider).

To understand "why" quality exists, culture and infrastructure need to be examined. These both need to be defined in the context of productivity, safety, and outcomes. Like software and hardware, culture and infrastructure interact and overlap. Culture helps to determine how information and knowledge flow. Information technology could be considered in this context or as part of infrastructure. Infrastructure is the "hardware" without which care would be impossible. It could include technology, physical plant, and equipment.

It is important to keep in mind that there are a number of different ways to "slice the pie" when it comes to dimensions of quality and that there is no one right answer. As more and more purchasers, payors, and others develop pay-for-performance initiatives, however, it is becoming clear that there is some convergence in the types of measures they are using. This convergence can also be attributed in part to the increasing sophistication in the area of health care quality measurement. As organizations like the

National Quality Forum continue to develop broad-reaching consensus on measures and measurement specifications, this trend is likely to continue.

For instance, in a 2004 survey, over ninety percent of pay-for-performance plans look at some type of clinical measures, and over half look at the use of information technology. The types of measure that tend to be used most often line up with most of the dimensions of quality we outline above – productivity, safety, outcomes, and infrastructure. One area that national efforts focus on but would not be appropriate for an HSCRC initiative is market share. The dimension of culture has received less focus in national efforts, but we feel is essential to track in Maryland to ensure that any changes are sustainable, despite the difficulties in measurement and change management.

Figure 5: Frequency of Use of Categories of Measures in Pay-for-Performance

Category of Measures in Survey	HSCRC Dimension	2003 n=28	2004 n=50
Clinical (e.g., HEDIS)	Productivity Outcomes	89%	94%
Patient Satisfaction	Outcomes	79%	30%
IT	Infrastructure	39%	56%
Efficiency	Productivity	57%	46%
Administrative/Market Share	n/a	54%	40%
Patient Safety	Safety	29%	13%
Other	Culture	n/a	n/a

Source: Med-Vantage, Inc. National P4P Survey. 2004. Presented at Second Annual National Health Information Technology Summit. Washington, DC. October 20, 2004.

One of the choices that the HSCRC will need to make is how much consideration to give to each dimension of quality. For instance, in Michigan, the Blue Cross Blue Shield of Michigan Participating Hospital Agreement Incentive Program seeks to promote the adoption of evidence-based clinical practices, improve the health of the community served, and reward hospitals for desired outcomes. To achieve these goals, it provides hospitals with incentives based on their performance in clinical quality (process), IT, community health. Clinical quality makes up half of a hospitals score, while IT contributes 40 percent and community health contribute 10 percent.[11]

The HSCRC should balance their vision, what can be measured well, and what should be measured in determining which dimensions to focus on first and their weight relative to one another. There is no question that each of these dimensions ought to be measured.

[11] Blue Cross Blue Shield and Blue Care Network of Michigan. www.bcbsm.com/providers/hospitals/hospital_enrollment_pha.html. Accessed last October 27, 2004.

Certainly, one of the loudest voices is that of the consumer in wanting to understand outcomes. But, to start with, unless a clear strategic imperative points otherwise, the weight assigned to these measures should be driven by our ability to measure well. This means that productivity, safety, and infrastructure will likely receive more focus initially until measurement methodologies are improved. The decision about how much weight will be assigned to each of these areas needs to be rooted in the strategic vision of the HSCRC in developing pay-for-performance.

Measurement Considerations

There are two main areas of considerations with regard to measurement: (1) measuring progress, and (2) measuring experience. A classic definition for measuring progress is to compare observed experiences with those that were expected to be observed based on previous local, regional, or national levels of performance. We will discuss how to measure experience later, with a specific focus on productivity and safety. However, first, we will look at how to capture what is expected the absence of pay-for-performance so progress can be determined.

Measuring Progress

Depending on the indicators selected, the difficulty in determining what is expected if no pay-for-performance system existed can vary. Three primary options exist for determining an expected trajectory: (1) baseline analysis, (2) external benchmarking, and (3) forecasting with assumptions.

The first possible approach – and the most straightforward – would be to capture baseline information on the selected measures. This will provide information on how much progress hospitals will make relative to the base period and thus the total impact of pay-for-performance However, it does not provide enough detail to understand the incremental impact of pay-for-performance (i.e., the impact of pay-for-performance relative to hospitals' individual quality improvement efforts). Moreover, this presents difficulties if the HSCRC decides to change and modify the indicators as the system evolves.

Another possible approach is to use an external comparison, such as to the US. This can be tricky as health care quality measurement is still a growing field and reliable and comparable national data may not exist. Moreover, an external comparison does not give a "pure" commentary on what would have been expected to happen in the absence of pay-for-performance. It is an imperfect control because more than one area of difference exists between Maryland and any external sample.

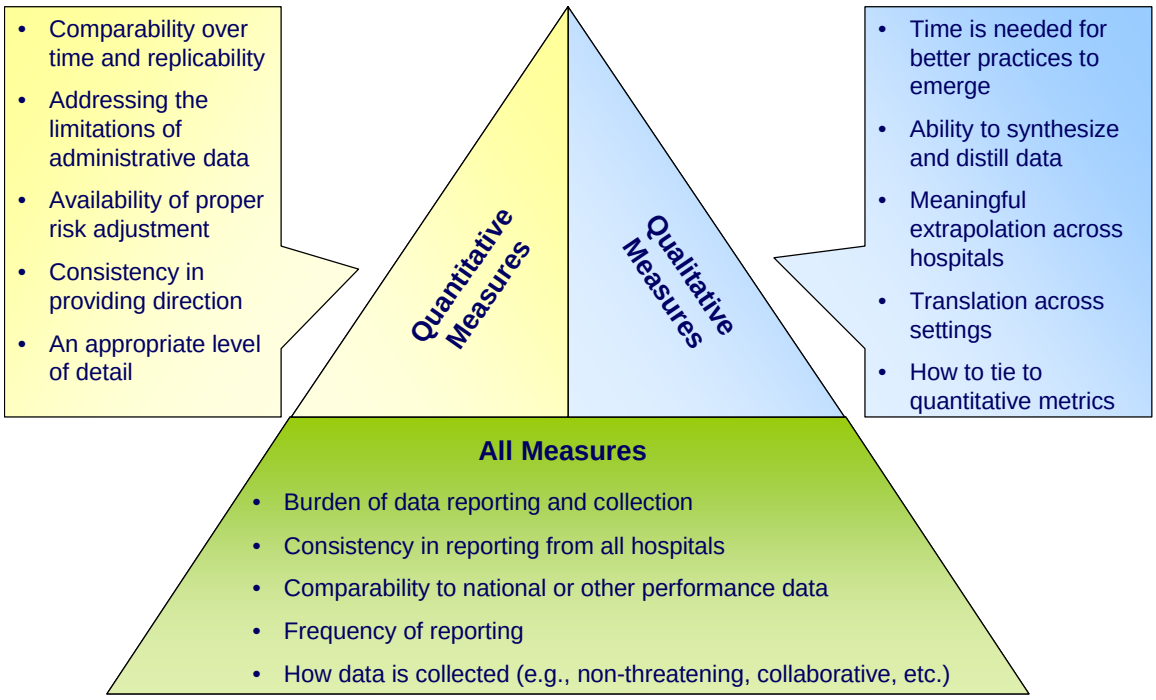
A third potential approach is to project what changes occur based on historical data or other trend assumptions. Again, this can present issues as historic patterns may not be the appropriate basis for predicting future behavior and other assumptions that could be used are also subject to question.

None of these approaches is perfect. But, experience in Maryland has shown that using a baseline type of approach with modifications for external experience and other assumptions may be the most sustainable way to proceed.

Measuring Experience

To understand how much progress is being made, actual performance will need to be measured. Different types of measurement will need to be employed to account for changes for each of these five dimensions of quality. Productivity, safety, and outcomes tend to be quantitatively measured while culture and infrastructure tend to be described in qualitative terms. We will look at each type of measurement and discuss the key aspects that can be used as criteria for evaluating data sets in the measurement selection process.

Figure 6: Major Measurement Considerations



Quantitative measurement specifically carries with it some intrinsic difficulties but provides the advantages of comparability and replicability. Measures need to be constructed in a way that outputs provide direction and instruction on where change needs to happen and how.[12] One of the issues raised repeatedly when looking at health outcomes quantitatively, proper risk-adjustment is essential to ensure valid comparisons.

[12] Thor J, Herrlin B, Wittlov K, Skar J, Brommels M, Svensson O. Getting going together: can clinical teams and managers collaborate to identify programs and initiate improvement? Qual Manag Health Care. 2004 Apr-Jue: 13(2): 130-42.

Another issue is the use of administrative data as a source of information for quantitative measurement. In the interest of consistency and efficiency, measures are not typically calculated from medical charts and records, but rather calculated from administrative data, which has limitations and biases. Research has shown that administrative data alone sometimes can omit information that whether a patient received a test or procedure when a review of the chart or record shows that the patient did receive that test or procedure. [13] For example, if a physician provides a patient with a sample medication in the emergency department rather than writing a prescription, this action does not get noted in administrative data but may affect how that process is measured. Supplementing administrative data with medical records can provide a richer set of information from which to understand a patient's experience, but can be a prohibitively burdensome endeavor.

Qualitative measurement too has its difficulties. It is a process that requires time and resources for data collection and synthesis. Better practices are not easily distilled from the experiences of many hospitals. They may or may not be translatable to different hospitals or settings. In order for the lessons to be drawn from this type of data in an effective and efficient manner, the data collection process should be focused and all stakeholders should be engaged in a collaborative process to discuss findings.

This qualitative data would be geared towards understanding the "why" beneath the "what" illustrated by the quantitative data. A potential consideration in looking at both quantitative and qualitative data in combination is sequencing of the data collection process. For example, quantitative data collected in the first period of measurement (e.g., year one) could trigger a series of qualitative questions in the second period of measurement (e.g., year two). Another potential approach would be to structure a semi-annual hospital survey as a follow-up to the quantitative data collection.

Regardless of what type of measurement is adopted, several other concepts need to be taken into consideration when determining how to measure actual experience, including the burden of collection, frequency of reporting, and manner in which data is collected.

How burdensome the data collection will be is a key consideration. Maryland has one of the most robust and accurate hospital data sets across the nation, but the ability to maintain such data requires a great deal of effort on the part of hospitals and also HSCRC staff. The amount of data, ease with which it can be collected, and the frequency with which reporting requirements (i.e., changes in measures and changes in measure specification) are made, can all increase the amount of effort required to develop and maintain the data for pay-for-performance. Another key consideration is how to ensure consistency across Maryland hospitals (for quantitative measures) or ensure the ability to extrapolate experiences across Maryland hospitals (for qualitative measures). The HSCRC may also want to think about if and how measures adopted in Maryland can be compared to the US so that Maryland's progress vis-à-vis the nation is not lost.

[13] Keating NL, Landrum MB, Landon BE, Ayanian JZ, Borbas C, Guadagnoli E. Measuring the quality of diabetes care using administrative data: is there bias? Health Serv Res. 2003 Dec: 38(6pt 1): 1529-45.

The frequency of reporting also influences how quantitative and qualitative measures can be used. Real-time data updates can be used to change processes in real time. For instance, making up-to-date patient waiting times available has been shown to reduce the amount of time needed to assign a patient to a bed and in turn reduce the amount of time that a hospital diverted patients from its emergency department.[14] If reporting were done annually, mathematical issues such as if and how to average, when to capture a point in time, how to synthesize, and how to account for outliers and variation, would need to be addressed.

Finally, the way in which measures are collected and shared can also influence their ability to create and/or facilitate change. A non-threatening mode of engagement is essential to continuous change and improvement.[15]

Measuring Productivity and Safety

The degree to which current data sets and better practice documentation can be drawn upon to facilitate pay-for-performance will influence the speed with which the payment system can accommodate these incentives and rewards. Of our five dimensions of quality, productivity is the one that has garnered the most attention in health services research. Historically, many different stakeholders have had clinical and financial interest in improving productivity. Safety has more recently become an area of focus, with the IOM drawing attention to safety as a critical issue for quality improvement. We will explore some of the particular considerations related to these two dimensions.

Figure 7: Additional Measurement Considerations for Productivity and Safety

Issues To Consider When Selecting Measures for Productivity and Safety
1. Whether the measures look at standard, routine or non-routine processes (i.e., volume and predictability of result)
2. Whether the measures capture the appropriate statistics given what kind of process they represent
3. How the measurement methodology can adjust for changes in volume given extrinsic forces like benefit design and structure of care provider
4. How the measures tie to existing clinical guidelines and protocols
5. The ability of the measures to adapt to changes in clinical guidelines and protocols

[14] Jamby L, Fraser SW. Using patient waiting time data to improve the hospital bed-assignment process. Jl Comm J Qual Saf. 2004 Jan: 31(1): 42-6.

[15] Dijkstra RF, Braspenning JC, Huijsmans Z, Peters S, van Ballegooie E, ten Have P, Casparie AF, Grol RP. Use of continuous quality improvement to increase use of process measures in patients undergoing coronary artery bypass graft surgery: a randomized controlled trial. JAMA. 2003 Jul: 2(290): 49-56.

When looking at productivity and safety, one invariably focuses on the process of health care. How efficiently are they performed? How effectively are they performed? How predictable are their outcomes? Volume and frequency are often viewed as threshold questions, as is the predictability of a process in resulting in a given output or outcome. Approaches taken in quality assurance more broadly often focus on standardization for high volume and high frequency processes. The wide variation in practice patterns in health care is well-documented,[16] but how often a process is performed as well as the intrinsic uncertainty associated with that process will influence how you want to measure that process and what you will be looking to change (i.e., standardize) based on the resulting measurements.

A useful construct is to divide health care processes into standard, routine, and non-routine. What you would measure to capture quality in these three different types of processes would differ. For instance, with a standard process, you would expect little variation in output and method so deviation from target output could be measured. With a routine process, little variation may be expected in method but perhaps some deviation in output (i.e., less predictability of outcome), hence deviation from a target would be inappropriate. Measuring the frequency with which the proper procedure (or the error rate for that procedure) is followed may be more appropriate. Non-routine processes are those that are most difficult to gauge broadly in the context of quality measurement. Here, what is important is interpretation and failures of interpretation.[17]

In addition to the frequency with which processes occur and the predictability of their output, benefit design and structure of the care provider can influence application of process. Medicare managed care has been studied to asses the impact of how plan design and coverage incentives can influence the receipt of services. One study looked at flu and pneumonia vaccination and smoking cessation counseling and concluded that the managed care model resulted in higher rates of this type of preventive service use.[18] Care for chronic conditions such as diabetes and asthma have also been studied to understand how benefit design, particularly around structuring self-managed care, can influence the process of care.[19] Health insurers are more actively marketing disease management programs for these types of chronic conditions, even including premium discounts for large customers who disease management in their benefits, such as with Aetna's Healthy Outlook Program. These issues become more important as more of health care costs become shifted to the consumer via higher deductibles and consumer-driven health plans and greater experimentation with different organizational models evolve.

The structure of the care provider can also influence the process and how quickly change can occur. This is another area where the treatment of chronic conditions provides a

[16] Wennberg JE. Dealing with Medical Practice Variations: A Proposal for Action. Health Affairs. 1984: 3(2): 6-32.
[17] Lillrank P, Liukko M. Standarad, routine, and non-routine processes in health care. Int J Health Care Qual Assur Inc Leadersh Health Serv. 2004; 17(1): 39-46.
[18] Landon BE, Zaslavsky AM, Bernard SL, Cioffi MJ, Cleary PD. Comparison of performance of traditional Medicare vs. Medicare managed care. JAMA. 2004. Apr 14; 291(14): 1744-52.
[19] Glasgow RE, Davis CL, Funnell MM, Beck A. Implementing practical interventions to support chronic illness self-management. Jt Comm J Qual Saf. 2003 Nov: 29(11): 563-74.

good basis for evaluation. One study of diabetes in the Netherlands demonstrated that use of diabetic specialist nurses (compared to internists) was linked to a much higher rate of adherence to recommended care.[20] Another recent study of diabetes care found that, in a for-profit setting, more tightly managed organizational structures (e.g., group or network models) tend to deliver recommended services more often than counterparts in less tightly managed structures. Interestingly, in a not-for-profit setting, no one management model proved to be easier to influence (i.e., ensure that they provided the recommended services for diabetes care) than another.[21]

In considering types of measures for productivity and safety, it is important to look at how they link to existing guidelines and protocols. For example, in the treatment of depression, researchers have shown recently that use, dosage, and duration of antidepressants according to guidelines can influence the likelihood of psychiatric hospitalization. Specifically, those patients on antidepressants with appropriate duration were less likely to be hospitalized than those who did not use the medication for the appropriate length of time.[22]

But adoption of guidelines and protocols takes time, even for chronic conditions that are often the focus of study and standardization. Duplication often exists and wording sometimes can be conflicting. From a measurement standpoint, this leads to tremendous difficulty in reaching agreement on numerators and denominators. From a process of care standpoint, this confusion can lead to variation in how the guidelines and protocols are adopted in care processes and algorithms. Leadership within the provider community is critical to adopting these guidelines and working towards uniformity. However, even once such uniformity is achieved, the impact of this does not become immediately apparent in quality metrics. For instance, the state of Delaware saw no significant change in quality indicators in the year before and the year after they implemented uniform clinical guidelines for diabetes care, although some minor improvements were seen in flu vaccinations.[23]

The measurement of productivity and safety has received a great deal of attention by health services researchers. The wealth of knowledge on this topic can inform the measure selection process and bolsters the case for focusing on these two dimensions in combination with infrastructure in the initial stage of a pay-for-performance system. How well a dimension can be measured should be one of the major driving forces in determining when it is folded in to measurement to support pay-for-performance.

---

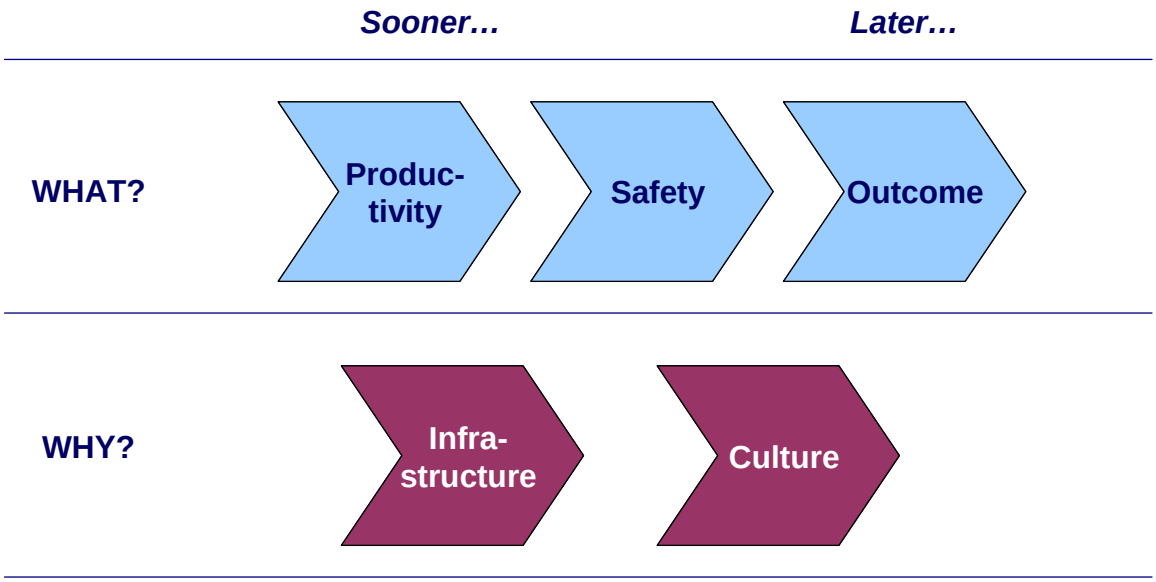
[20] Dijkstra RF, Braspenning JC, Huijsmans Z, Peters S, van Ballegooie E, ten Have P, Casparie AF, Grol RP. Patients and nurses determine variation in adherence to guidelines at Dutch hospitals more than internists or settings. Diabet Med. 2004 June: 21(6): 586-91.

[21] Kim C, Williamson DF, Mangione CM, Safford MM, Selby JV, Marrero DG, Curb JD, Thompson TJ, Venkat Narayan KM, Herman WH. Managed care organization and the quality of diabetes care: the Translating Research Into Action for Diabetes (TRIAD) Study. Diabetes Care. 2004 Jul 27 (7): 1529-34.

[22] Charbonneau A, Rosen AK, Owen RR, Spiro A 3rd, Ash AS, Miller DR, Kazis L, Kader B, Cunningham F, Berlowitz DR. Monitoring depression care: in search of an accurate quality indicator. Med Care. 2004 June: 42(6): 522-31.

[23] Gill JM, DiPrinzio MJ. The Medical Society of Delaware's Uniform Clinical Guidelines for diabestes: did they have a positive impact on diabetes care? Del Med J. 2004 Mar; 76(3): 111-22.

Figure 8: Phasing In Dimensions Based on Ability to Measure



Payment Considerations

At the beginning of this paper, we reviewed the three proposed streams of funding. Five other major design considerations are apparent at this time. No doubt as the design process begins and becomes more detailed, additional items will need to be considered in terms of direction and implementation.

Figure 9: Major Payment Considerations At This Time

Major Questions to Address in Design of Payments
1. Dimensions of Quality: All or some and in what order?
2. Competitive or Not: Should penalties be used?
3. Thresholds: Where should the "bar" be set?
4. Pilot: Limit the number of hospitals, measures, payments, or all of the above?
5. The Fragility of the Safety Net: Increasing divergence of "haves" and "have nots?"

First, what dimensions should pay-for-performance focus on? The Steering Committee recommended that a composite score be created. By doing so, a mechanism for weighting the dimensions relative to one another will be created. The scoring process will also help with comparability over time as different measures are brought into the system. However, the question of which dimensions to include first remains. As was

mentioned above, the ability to measure well should drive this decision. Productivity and safety are the two dimensions for which the most literature exists. Infrastructure is a pre-cursor to improving performance and is the foundation of sustained improvements.

Second, will the system be competitive or non-competitive? And, relatedly, will participation be voluntary or mandatory? Hospitals will be very interested in the revenue implications of pay-for-performance in addition to the quality improvement implications. In a competitive system, hospitals would compete against each other in terms of performance and penalties may be applied. In a non-competitive system, hospitals would be held accountable for progress from a baseline level. Any competitive system would have to carefully ensure that comparisons are valid, perhaps based on the characteristics of quality that a set of hospitals cannot change. Even in a non-competitive system, some degree of comparison across hospitals may be required. The related question of whether the system should be voluntary or mandatory is a tricky one. The HSCRC should strive to minimize the amount of adverse selection due to pay-for-performance.

Third, where should the "bar" be set when setting thresholds for rewards? This is a complex and controversial topic. Thresholds may vary by indicator and will also change over time. Certainly, external benchmarks for different indicators exist and are being developed, though many are proprietary. The option also exists of using Maryland-specific experience or targets. The controversy arises when characteristics a hospital cannot change (e.g., location) affect its ability to meet that threshold and exceed it.

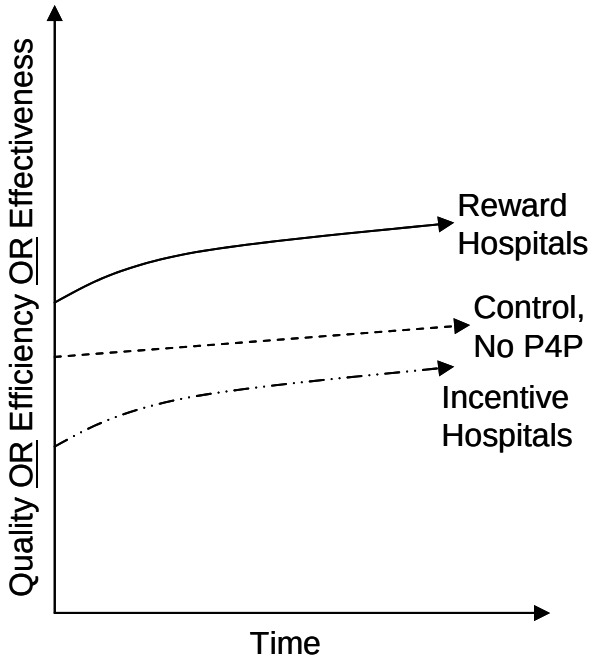
Fourth, how could the HSCRC structure a pilot to test the impact of the payment? It could be limited to a few hospitals or be spread across all hospitals. It could test a subset of measures or all of them. It could test only rewards, only incentives, or a limited amount of both. Piloting a proposed approach with a few hospitals instead of all of them will allow for a more rapid pilot as well as the ability to test more. Testing a subset of measures provides a sense of the complexity of the data collection process. Testing anything other than both rewards and incentives at once does not effectively allow for an understanding of how the two can interact to push hospitals to improve. Conducting pilot project is the key to success.

Fifth, will pay-for-performance lead to increasing divergence in "haves" and "have nots" and threaten an already fragile safety net? The hope and intention is that this will not occur. In structuring the combination of incentive and reward payments, several scenarios are possible and providing a composite score (as recommended by the Steering Committee) rather than incentive and reward payments for individual measures can lead to some complications.

Three scenarios should be considered in understanding how rewards and incentives can affect the performance gap between hospitals. To develop these scenarios, we assumes that hospitals who would receive incentive payments are not performing as well as the average hospital. The critical event to look for in these scenarios is when the hospitals receiving incentive payments "catch up" to the average hospital without pay-for-performance and the rewards hospitals.

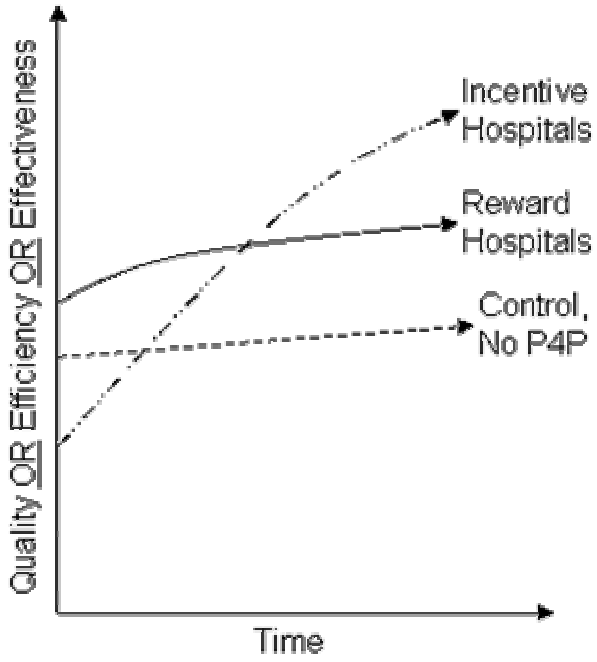
In the Scenario 1, the hospitals receiving incentive payments do not improve at a rate fast enough to catch up with the average hospital that is not being paid for performance. Figure 10 shows how the lines never meet. Here, incentives do not drive change substantially more than the normal un-systematic synthesis of knowledge from the field. The performance of hospitals receiving incentives would plateau. This means that, at best, pay-for-performance kept the gap between the "haves" and "have nots" from widening but did not make dramatic improvements in closing that gap. Here, the level of payment and the way in which it links to what is measured should be re-evaluated.

Figure 10: Scenario 1 – "Ongoing Improvement"



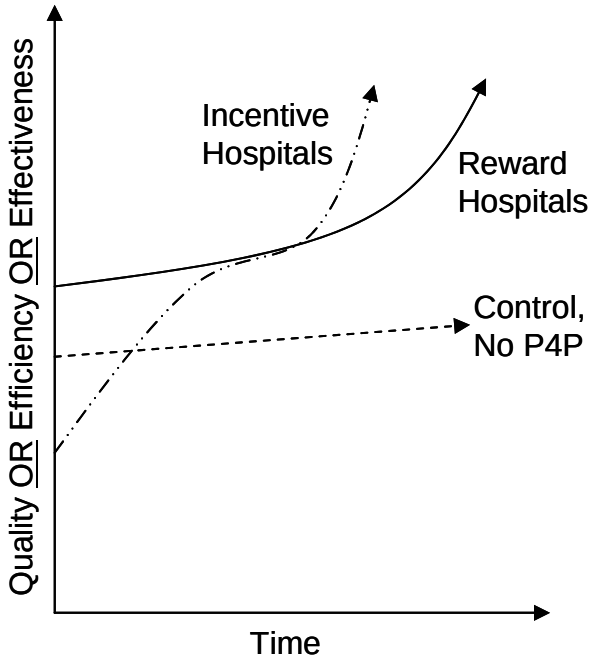
Scenario 2 shows a more positive result. In this case, the hospitals receiving incentive payments would improve a rate faster than those receiving reward payments and average hospitals without pay-for-performance. Figure 11 shows how the performance of incentive hospitals surpasses that of reward hospitals. Here overall quality will improve over time, but hospitals receiving incentive payments will improve more rapidly, thus potentially leading to a reversal of the "haves" and "have nots." If this occurs, the reversal will likely be small and temporary but long enough to prompt hospitals receiving reward payments to pick up the pace of their improvement efforts. This type of scenario is one that pay-for-performance should seek to foster.

Figure 11: Scenario 2 – "Continued Improvement"



When payments are well structured, there will be instances where improvement with those receiving incentives will parallel improvement with those receiving rewards, but there will also be instances where incentive hospitals out-perform reward hospitals and vice versa. Over time, this will lead to continued improvement so that the average level of performance after implementation will be higher than that when it starts. However, unless there are major medical advances, no dramatic improvement will occur. Scenario Three and Figure 12 shows how this would cause an inflexion point in hospitals' improvement if the payments encouraged hospitals to embrace the breakthrough. At this point, the pace of change increases markedly for all hospitals. This would signal the success of pay-for-performance in systematically disseminating new knowledge.

Figure 12: Scenario 3 – "Breakthrough"



How to Compare Hospitals

The concept of comparison is important because it dictates who gets what type and what level of payment and how the HSCRSC can monitor progress over time. Hospitals will be compared both on the basis of "what" their performance and quality is and "why" that level exists. By pairing these comparisons, a feedback look is created. As hospitals learn more about "why" their performance is at a certain level, they have more information to improve "what" their performance is.

How comparison groups will be defined and structured and then how payments will be assigned are the two key questions to explore in designing a comparison framework and model. We highlight three elements – hospital characteristics, dimensions of quality, and types of payments – as factors to consider in answering these two questions and building this framework. The degree to which each of these is used should be determined by the HSCRC using criteria such as

- How complex the comparison process would become;
- How the pay-for-performance comparison process interacts with other payment system methodologies (e.g., ICC, ROC, etc.); and
- How the process can be explained to consumers and others as a means to further promote accountability and improve quality via the payment system.

Hospital characteristics are one important element to consider in creating a comparison model. One can build on the idea of grouping hospitals with like peers, an idea that that the HSCRC currently uses to evaluate hospitals' Reasonableness of Charges (ROC). For instance, groups for pay-for-performance purposes can be created based on quality-related characteristics or criteria. Individual hospitals would then receive reward or incentive payments based on their performance relative to their group. To create these groups, it may be beneficial to look at what characteristics a hospital cannot change with regard to quality. By grouping hospitals this way, it is possible to compare hospitals on those characteristics they will be expected to change and evaluate the effects of rewards and incentives in helping them do so. As hospitals improve quality, the characteristics that hospitals were expected to change will move into a different category – characteristics that they will have already changed, and the expectations for change system-wide will need to be updated. In other words, what hospitals are expected to change (i.e., what hospitals are compared on) changes as improvement takes hold.

In addition to hospital characteristics, the dimensions evaluated need to be taken into consideration when comparing hospitals. Unless thresholds for reward payments and criteria for incentive payments are set for each individual measure, some aggregating methodology needs to be created to determine the basis on which a hospital receives a reward, an incentive, or nothing at all. For example, each hospital can receive one overall quality score from which a decision on payment is made. However, the work by CMS with its demonstration with Premier would suggest that it would be more prudent to set thresholds by at a lower level of detail. Looking at scores for individual measures

could be a prohibitively burdensome process while looking at one score for all measures may mask important information about the progress being made in any specific area.

The HSCRC needs to make a choice on how specific the Maryland system will be and how that specificity will be developed. For example, should Maryland's system provide payments based on performance within dimensions so that hospitals would have their performance for patient safety assessed differently than performance on productivity? If so, how does the composite scoring methodology get developed?

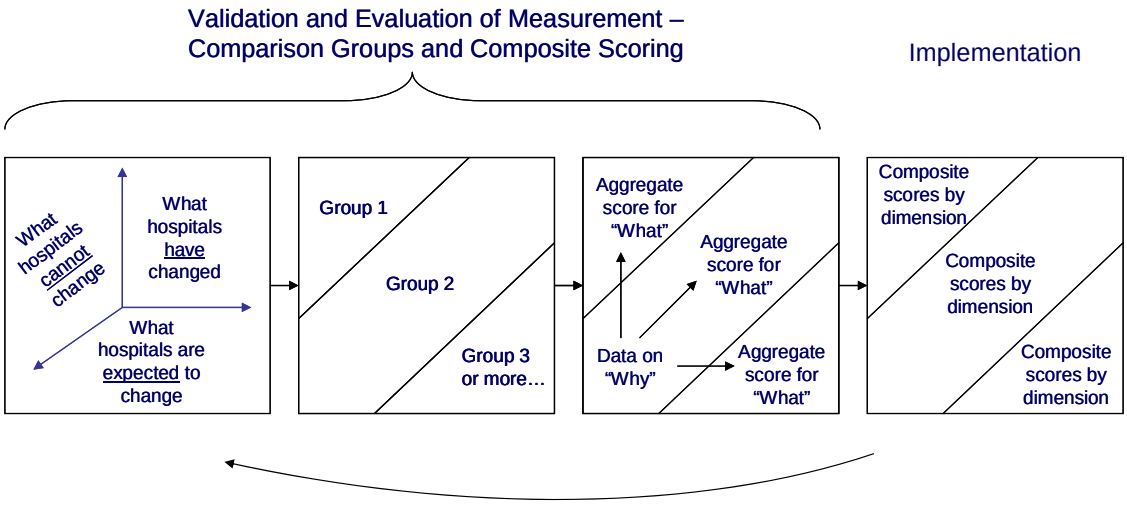
These decisions need to be rooted in philosophical and practical matters. Philosophically, the argument can be made that making comparisons by dimension may not adequately provide a way to look at total quality of care and may be difficult for consumers to decipher. Practically, the key point is revolves around the HSCRC's Steering Committee's recommendation to use composite scoring and making sure that the scoring can be done at the appropriate level of analysis.

Moreover, the decision is not one that, once made, remains static. This is an evolutionary process. Once groups are determined, changes in what expectations are for improvement can alter how the groups should be organized. The composite scoring process needs to start first by looking at information on why hospitals are at different performance levels and apply that to aggregate data on the dimensions that measure "what" (e.g., productivity, safety, and outcomes) to develop a dimension-specific composite scoring methodology.

Finally, the nature of the payments received by hospitals can help the HSCRC in understanding what impact the reward and incentives are having. Although this has no ability to help the HSCRC determine who receives payments and rewards, it is an important exercise to evaluate the historical effectiveness of the types and levels of payments. Hospitals receiving rewards should be first compared with each other before they are compared with hospitals receiving incentives and vice versa. This provides a sense of the variation inherent to the group of hospitals receiving rewards or incentives. With this, the assessment of differences in performance between the reward-based hospitals and the incentive-based hospitals will be more accurate. This comparison then provides the HSCRC with further direction on how to structure their payments.

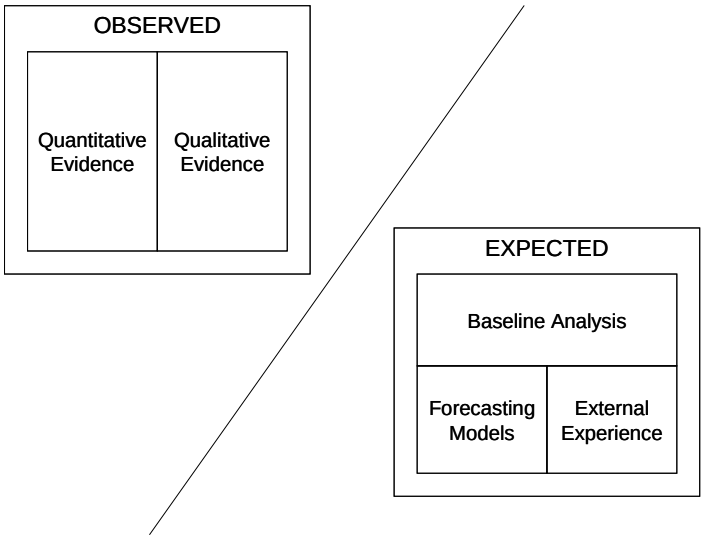
This discussion about how to compare hospitals so far has focused on how to do so with quantitative metrics. Qualitative methods are more appropriate to explore why groups of hospitals demonstrate different patterns of performance. This information will be valuable in determining how to construct composite scores. In addition, interview-based and observation-based qualitative analysis will better highlight for hospitals what emerging better practices are so that they might benefit from the experience of others. This compendium of "better practices" can supplement the "report card" type of quantitative data hospitals can use to improve quality. Both quantitative and qualitative data collection methodologies are needed to paint the picture of hospitals' performance in full color.

Figure 13: A Framework for Comparing Hospitals



With this complete picture, the HSCRC can then assess how hospitals are doing vis-à-vis hospitals with similar characteristics, how hospitals are doing within specific dimensions or quality or medical conditions, and the impact of rewards relative to incentive payments. In tandem, the HSCRC also needs to use some objective standard against which to measure how well hospitals across the state are doing overall to determine progress. This objective standard could be developed using baseline analysis, external benchmarking, forecasting assumptions, or some combination of the three as discussed in the section on *Measuring Progress* but needs to tell the HSCRC what the picture in Maryland would be in the absence of pay-for-performance.

Figure 14: A Way to Measure Progress



Comparing hospitals so that the HSCRC might pay them for performance is an on-going and complex process. Factors such as hospital characteristics, types of payment, and dimensions of quality all enter into the equation of how to create comparison groups. Composite scoring will need to be tailored to the groups and qualitative data will need to be considered. The ability for the HSCRC to communicate to the public and its constituencies how this comparison occurs is critical to its decisions about how to set up these comparisons, as is the ability of the HSCRC to maintain a complex comparison process in an already complex payment system.

Possible Limitations

Direct Care Provider Involvement and Support
In laying out a framework, we have focused on organizational changes but the role of direct care providers deserves special attention. They need to buy in to the philosophy linking performance to payment in order to for the initiative to succeed.

While HSCRC may not have direct influence on these direct care providers, it can create a structure (e.g., a "reach out and listen" program) to reach out to them on a regular basis so that they can provide feedback during the pilot phase and beyond. Direct care providers should also be involved in the design phase in an advisory capacity.

Furthermore, the HSCRC can appeal to direct care providers by providing the funding for infrastructure that can improve their ability to perform their clinical work. As we discussed earlier, the first step in paying for performance is paying for the necessary people, processes, and technologies required to perform. Some direct care providers may not be able to access this required infrastructure because of their hospitals' financial condition, the public good nature of these investments (i.e., uncertain financial ROI), or other reasons.

The HSCRC needs to demonstrate early on that this effort is not about penalizing clinicians but is about working with them to help hospitals develop better ways to provide health care. Providing direct care providers with technology and infrastructure to help them improve their ability to provide high quality health care is a first step.

Accommodating Patient and Community Preferences
Not only is it important to consider the direct care provider's role in developing and sustaining a pay-for-performance system, it is important to look at the role patients and their communities can play.

They are an important lever for change, advocating for progress as health care costs increase and are passed through to them via higher health insurance premiums, higher rates of uninsurance, and worsening community health as appropriate care is not sought. The broader public interest in accountability makes it important for the HSCRC to be deliberate in how it chooses to share data and results about pay-for-performance.

From a measurement and comparison standpoint, though, differences in patients and their preferences could lead to some variation in how care is performed and received at a hospital. For example, a patient may reject a tetanus vaccination despite coming to an emergency room with a deep wound caused by an old rusty nail due to a negative family experience. Or, a patient may reject recommended medical care due to religious beliefs. Some flexibility in the comparison process – perhaps having ranges instead of hard targets for thresholds or sustained improvement – is needed to accommodate for patient variation.

Rapidly Changing Technology
Robust quantitative measurement tools should have the capacity to be modified with new knowledge. However, technology may progress at a faster pace than the measurement tools can accommodate. This can occur both with clinical and information technology. In the case of clinical technology, when measurement cannot keep pace with change, the comparison process needs to account for any changes appropriately.

Information technology is more complex. Changes in information technology will change the data that is available, but, more importantly, revolutionize the way care is provided. For example, there is a movement to move towards healthcare information exchange and interoperability (HIEI). With HIEI, direct care providers will have access to information that they may not have before and be able to access that information in a time frame that enables it to be used in clinical decision-making.

This is revolutionary, not evolutionary, change and the answer is still not clear as to how measurement can keep up with this. The HSCRC will need to draw on expertise in the field to address these measurement issues as they arise. However, it is clear that pay-for-performance should promote this type of change.

The Impact to Patient Care, Hospitals, and the Payment System

How will this affect the health of Marylanders? More specifically, how will it improve the health of Marylanders? This is the ultimate question that pay-for-performance seeks to address. The answer is not straightforward.

Without pay-for-performance, quality improvement will likely occur slowly, in an uneven fashion across the state. A pay-for-performance initiative would to help the best of the best improve more quickly as well as help even out the rate of progress statewide, ensuring that all Marylanders are able to access high quality care.

This comes about via a feedback loop in which data on specific patient encounters is captured in and by hospitals and then returned to them in the form of payments to encourage certain behaviors and the information to help them do so. Hospitals continually work to identify what they can do to improve quality within their walls, and having the HSCRC data on how they compare to others ("scorecard") as well as what better practices are gives them another source of data from which to draw ideas.

Figure 15: Payment System Impact

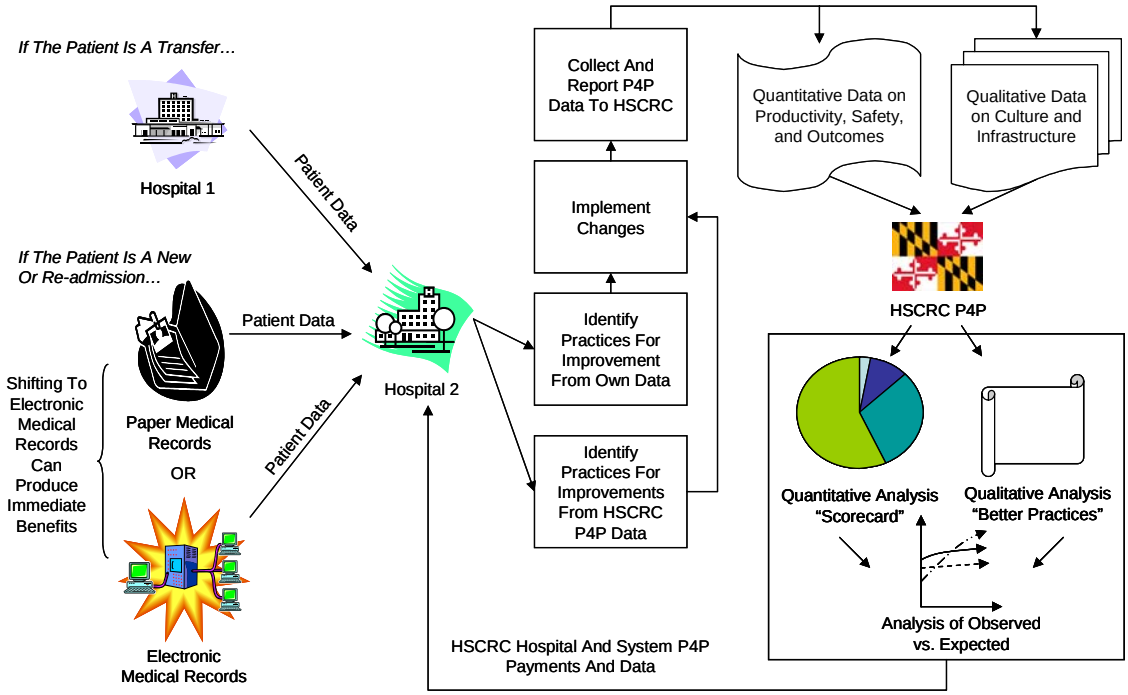
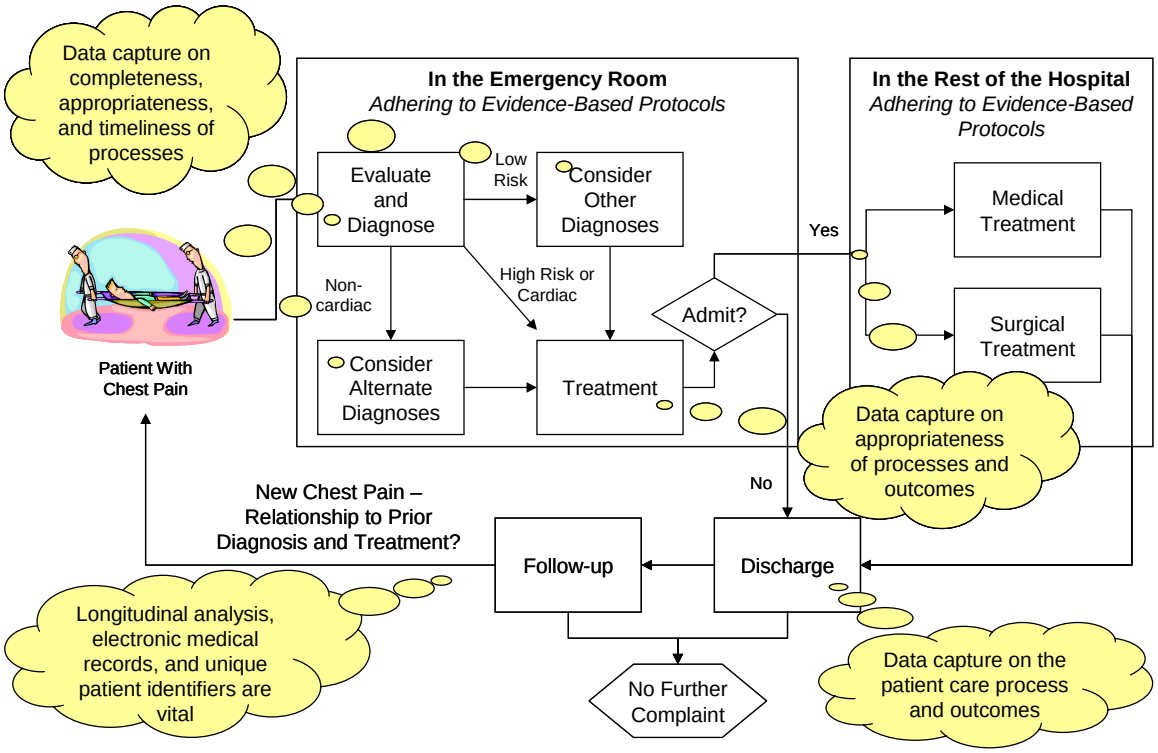


Figure 16: Pay for Performance and Patient Care



In terms of patient care, a similar feedback mechanism is created where a hospital can begin to understand how well it is doing in terms of clinical processes such as evaluation, treatment, and discharge. At each point during the care process, critical information is captured on completeness, appropriateness, timeliness, and outcomes. Much of this information may be captured already in hospitals' individual quality monitoring.

The key here, and what makes the pay-for-performance feedback loop distinct from hospitals' current quality monitoring feedback loops, is that pay-for-performance data have revenue implications for hospitals. The revenue impact increases hospitals' focus on the quality tracked by the measures and its efforts to ensure that improvement occurs in each patient encounter. The additional revenue available to hospitals can also facilitate their adoption of such key infrastructural enhancements, such as healthcare information exchange and interoperability (HIEI).

HIEI is particularly important since its benefit, when adopted by a large number of hospitals, exceeds the sum of the benefit accrued by each of those hospitals. Also, those benefits do not necessarily rest with the hospital alone – payors in particular benefit greatly from the improved information exchange, as do physician offices, pharmacies,

laboratories, and other stakeholders.[24] The return on these investments needs to be measured in more than just dollars, but in the improvement of quality over time.

For example, if a patient were to come into a hospital's emergency department with chest pain, an effort is first made to evaluate and diagnose the patient. If the hospital and other providers had HIEI, this would allow them to access critical information on the patient's history, prior hospitalizations and prior treatment. This could allow caregivers to make informed decisions about diagnostic and treatment procedures, including not repeating a test or avoiding an adverse drug event.

In that same example of the chest pain patient, during the process of evaluation and diagnosis, valuable information can be captured on how quickly the team is responding, or how appropriately and completely it is doing so. Perhaps the patient is a cardiac patient, or one at high risk. Alternatively, the chest pain may not be cardiac or the patient is a low-risk patient. In this case, alternate and additional diagnoses would be considered. Once the patient gets treated in the emergency department, a decision is made whether to admit the patient for further medical or surgical treatment. Throughout the treatment processes, information on the appropriateness of care (e.g., adherence to evidence-based protocols) and outcomes can help paint a picture of how well the hospital is responding in this instance. When the patient is discharged, further information can be collected to evaluate the patient's experience as a whole (e.g., process, satisfaction, etc.).

For the payment system, this means continual improvement in efficiencies and quality. It means a focus on innovation and performance as well as increased accountability.

Conclusion

This a major undertaking on which the HSCRC has embarked. This paper has tried to distill the key recommendations from the HSCRC Steering Committee report and develop a framework from which a work group can then complete the design of the system, conduct an initial pilot, and move forward with implementation.

Lessons from other sectors, such as education, environmental management, government, and advertising can provide a good starting point. Their experiences with pay-for-performance indicate that using experts to set goals, collaborating with performers to identify indicators, considering outcome data with other quantitative and qualitative data, staying focused on continuous improvement, giving performers flexibility to achieve their own missions, and being open to considering changes in measurement are vital to creating a sustainable program.

Lessons from health care are more limited, but provide some clear signals on how to proceed. Quantitative data should be supplemented with qualitative data. The choice of measures and data collection needs to proceed carefully and should build on the large body of health care quality measurement that already exists. Criteria for quantitative and

[24] Middleton, B. Center for Information Technology Leadership, Boston, MA. Assessing Value/Calculating ROI. Presentation at the Second Annual National Health Information Summit. Washington DC, October 20, 2004.

qualitative measurement are clear. Dimensions of quality that should be addressed include productivity, outcomes, safety, culture and infrastructure – these areas are seeing greater emphasis and convergence as pay-for-performance programs proliferate. Our ability to measure well will drive our early efforts, meaning productivity, safety, and infrastructure will likely be areas of initial focus.

Some of the major principles on which to design payment mechanisms are also clear. Paying hospitals to ensure that the right people, processes, and technologies are in place (i.e., infrastructure payments) is a critical first step in paying for performance. Paying hospitals for meeting and exceeding targets could lead to increased disparities and reduced access to quality health care in Maryland hospitals if not coupled with paying hospitals for achieving sustained improvements.

The way to proceed with design, methodology, and testing also has some clear questions that need to be answered. To determine who gets rewards and who gets incentives, a composite scoring system should be created for dimensions of quality or for specific medical conditions. The implications for this composite scoring specificity to medical conditions or dimensions of quality is that an overall hospital performance composite score may not always be possible to construct by summing up the scores from the dimensions of quality of specific medical conditions. For example, a composite scoring for surgical procedures may be constructed by summing up the scores of individual surgical procedures measured under the pay-for-performance initiatives and adjusted according to patient, hospital, and even provider variables. Similarly, a composite score may be constructed for process of care indicators such as productivity and safety. However, there composite scores may not automatically add up to a grand composite score representing the hospitals overall performance. Although this recommendation is both conceptual and "mechanical" in nature, its true implications will only be evaluated once the measures are chosen and pilot tested. Consequently, these composite scores should be compared in a way that hospitals perceive as fair, such as in comparison groups determined by characteristics they cannot change about their organizations. Evaluating whether the rewards and incentives are at the right levels and right proportions relative to each other requires looking at their individual impact as well as the difference between the two types of payments. Finally, direct care providers need to be brought in to the process. Without them, sustainable improvements cannot occur.

RECOMMENDATIONS

Recommendation 1: Decision on Terminology

Throughout the report, it was argued that the initial phases of the HSCRC initiative will best measure "performance" of hospitals via indicators. The term "pay-for-performance" could be considered to be replaced by "quality-based reimbursement" if a decision is made in the next phases of this initiative about the ability of the indicators to directly measure quality.

Recommendation 2: The Structure and Composition of the Evaluation Workgroup

The role of the Evaluation Workgroup, as defined in the HSCRC report, is crucial for the success of this initiative. The conceptual framework presented in the present report as well as the indicators to be chosen for pilot testing will be among the first tasks of the Evaluation Workgroup. Therefore, the expertise represented on the Group and the frequency of its deliberations have to be organized and coordinated diligently.

Recommendation 3: Prioritization of Focus Areas

It is recommended that the HSCRC should prioritize the measurement of the focus areas during the implementation of a pay-for-performance model. The focus areas are: productivity, hospital infrastructure, safety of care, and organizational culture/readiness for improvement in care delivery. Figure 17 shows the proposed sequence of the prioritization.

Figure 17. Prioritization of focus areas and the sequence of their phasing into a pay-for-performance model

<table>
<tr><th>Sooner…..</th><th></th><th>Later</th></tr>
<tr><td rowspan="2">Productivity and hospital infrastructure measures</td><td>Safety of care measures</td><td>Impact/outcomes measures</td></tr>
<tr><td colspan="2">Organizational culture and readiness measuring</td></tr>
</table>

Recommendation 4: Implementation Phases

The pay-for-performance initiative (or, alternatively, the quality-based reimbursement initiative) should have pilot testing phases to establish the relationships between the rewards and incentives allocation strategies. It is recommended that there should be five consecutive phases to achieve the above goal. The recommended phases are shown in Figure 18.

Figure 18: Future Phases and Activities

Phase 2:

1. Develop concurrence on dimensions with this paper as starting point
2. Finalize criteria for evaluating measure with this paper as starting point
3. Design and conduct alpha pilot project (small subset of hospitals)
    - Identify and pick specific measures (3-5 at most)
    - Develop composite scoring methodology at appropriate level of detail
    - Determine pilot incentives, rewards, and infrastructure payments
4. Evaluate alpha pilot results

Phase 3

5. Design and conduct beta pilot project (larger subset of hospitals)
6. Evaluate beta pilot results

Phase 4

7. Revise the measures, the composite scoring methodology, and their implications for incentives, rewards, and infrastructure payments
8. Design a full-scale implementation strategy for statewide implementation

Phase 5

9. Conduct full-scale implementation. A "transition" period will be built into this phase for all hospitals which were not in the alpha and beta pilot cycles.

Recommendation 5: Beyond Financial Factors

The HSCRC Report proposed that the level of funding for the incentives and the reward strategies would be based on indicators, cost-benefit analysis, and return on investment (ROI). It is recommended that those considerations go further by considering factors beyond financial terms and include such performance aspects as hospital compliance with better practices protocols, adherence to clinical guidelines, and demonstrated reductions in factors enhancing safety of care.

Recommendation 6: Quantitative and Qualitative Measurement Strategies

While indicators measure quantitatively aspects of performance in hospitals, it is recommended that quantitative measures should be supplemented by qualitative measurements. Productivity and safety measures lend themselves to such joint measurement approaches (quantitative and qualitative) since organizational culture, adoption of technology, communication channels, are among the factors that could best be measured qualitatively to supplement the quantitative data captured by specific indicators.

Recommendation 7: Hospital Cluster Analysis

When comparing hospitals, HSCRC should use a method to group similar hospitals to assess the changes in performance over time. Hospital characteristics, types of payment and dimensions of performance would be among the stratifying variables. Indirectly, such stratification and grouping of hospitals would provide an opportunity to also adjust the findings by patient characteristics.

Recommendation 8: Comparisons Using Baseline References

While incentive and reward-based strategies will be monitored during the pilot phases, it is necessary to have a benchmark or a baseline to show what performance would have been in the absence of these strategies. The baseline can be calculated from previous performance profiles of Maryland hospitals and supplemented by national data. Two corollaries to this recommendation are worth considering.

Corollary 1: The most sustainable way to proceed in determining the expected profile in improvement across the two strategies of incentive and reward systems is to compare and contrast the observed improvements with the baseline data. IN this instance, the baseline will become the "expected" while the observed improvements will be the "observed." The gap between the observed and the expected will constitute the first method of assessing the nature and extent of improvement.

Corollary 2: Paying hospitals for exceeding certain targets can lead to increased disparities if the payment system as rewards or incentives is not adjusted for the starting baseline performance levels of hospitals when sustained improvements are observed.

Recommendation 9: Direct Care Providers as a Target Group

Direct care providers should receive the support in technology and infrastructure to help them improve their ability to provide a high quality of care. Considering the needs of the direct care providers should be part of the strategies associated with either the reward or the incentive payment schemes.

Recommendation 10: Lessons from Other Industries

Lessons should be gleaned from other industries. There are generic models that have shown to be successful in numerous industries and are recommended to be beneficial to the HSCRC pay-for-performance goals as well. Among those generic models is the use of experts to set goals, ongoing collaboration with direct care providers to identify indicators of performance, the merging of quantitative and qualitative methods of analysis, the focus and promotion of continuous performance improvement, the provision of direct care providers with the tools and infrastructure to achieve their missions, and the built-in flexibility of the system to accommodate new knowledge when available from the field.