An Investigation into
CMS’s Value-Based Purchasing Proposal

By
Jake Ing

A Master’s Paper submitted to the faculty of
the University of North Carolina at Chapel Hill
in the partial fulfillment of the requirements for
the degree of Master of Public Health in
the Public Health Leadership Program.

Chapel Hill
2003

[Signature]
Sue Tashon-Rinehart, PhD, Advisor

[Signature]
Greg Randolph, MD MPH, Second Reader

24 June 2008
Date
“If a physician make a large incision with the operating knife and cure it... he shall receive ten shekels in money.

“If a physician make a large incision with the operating knife, and kill him... his hands shall be cut off.”

– Code of Hammurabi, 1870 BC
ABSTRACT

Tasked by the government through the Deficit Reduction Act (DRA) of 2005 to transform Medicare from a passive payer into an active purchaser of care, the Centers for Medicare and Medicaid Services (CMS) has been investigating methods of incorporating pay-for-performance (P4P) models into Medicare. When CMS published its Value-Based Purchasing (VBP) proposal for overhaul of Medicare inpatient reimbursements in November of 2007, the proponents and skeptics alike had much to say about the new plan. Some approved strongly. Some disapproved vehemently, while most others were somewhere in between.

Certain themes surrounding VBP continue to surface. What do stakeholders believe are the key issues surrounding P4P and the VBP proposal? How and why do payers and providers differ in their support for VBP? Do all stakeholders find common ground in the VBP, or are some aspects of VBP particularly agreeable or distasteful? Has CMS structured the VBP proposal in ways that will maximize VBP’s chances of success? Does VBP pose a threat to small and rural hospitals already struggling to stay afloat? Does the literature support CMS’s claim that VBP can improve quality of health care while reducing costs?

In order to address these questions, I performed a systematic review of the literature, reviewed media accounts, interviewed key stakeholders, and studied the stakeholder responses provided to the Senate Finance Committee Roundtable on Hospital Value-Based Purchasing. I found that payer groups were eager to introduce VBP, while the hospital and professional groups were concerned that VBP threatened their own interests, especially regarding budget neutrality. However, I also found some important items of consensus as well that could lay the groundwork for the future of VBP.
ACKNOWLEDGEMENTS

I would like to thank my advisor, Sue Tolleson-Rinehart, for generously providing her time and assistance in researching and writing this paper, as well as training me to acquire the technical skills I needed to complete this paper. I would like to thank my second reader, Greg Randolph, who also helped me formulate my ideas surrounding this topic. I would also like all my respondents, Jeffrey Spade, Nancy Foster, Brian Goldstein, and Thomas Valuck for their generous time and helpful responses and comments. I want to thank all my classmates who supported my research efforts and challenged my point of view on this topic with thought-provoking queries and suggestions. Finally, I would like to thank my family, Jack, Yvonne, Lenny, and Astrid, for their continued support of my educational goals.
# TABLE OF CONTENTS

ABSTRACT ................................................................................................................................... ii  
ACKNOWLEDGEMENTS ........................................................................................................ iii  
BACKGROUND ........................................................................................................................... 1  
  Introduction ............................................................................................................................ 1  
  The Medicare Modernization Act (MMA) of 2003 and Reporting Hospital Quality Data for Annual Payment Update (RHQDAPU) ................................................................. 2  
  Deficit Reduction Act (DRA) of 2005 ...................................................................................... 3  
  The 2007 update to RHQDAPU, and the HCAHPS survey .................................................. 4  
  The Premier Hospital Quality Incentive Demonstration (HQID) Project .......................... 5  
  CMS’s proposed plan to implement inpatient P4P for Medicare: Value-Based Purchasing ................................................................................................................................. 5  
  The Performance Assessment Model and Measure Development/Selection ....................... 6  
  Phased transition from RHQDAPU to VBP ........................................................................... 8  
  Funding of incentive payments in a budget-neutral fashion.................................................. 8  
  Supporting research and hospital feedback ........................................................................... 9  
  Search Strategy .................................................................................................................... 11  
  Premier HQID studies .......................................................................................................... 11  
  Other P4P papers ............................................................................................................. 22  

THE SENATE FINANCE COMMITTEE ROUNDTABLE ON HOSPITAL VALUE-BASED PURCHASING ............................................................................................................. 26  
  Overview of the Senate Finance Committee Roundtable on Hospital Value-Based Purchasing ......................................................................................................................... 26  
  Performance Standards ........................................................................................................ 27  
  Quality Measures ................................................................................................................ 27  
  Structure of Incentives ........................................................................................................ 29  
  Implementation .................................................................................................................. 30
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where They Sit: Advocacy</td>
<td>31</td>
</tr>
<tr>
<td>Where They Stand: Support for VBP</td>
<td>33</td>
</tr>
<tr>
<td>STAKEHOLDER RESPONSES</td>
<td>38</td>
</tr>
<tr>
<td>Rationale</td>
<td>38</td>
</tr>
<tr>
<td>Methods</td>
<td>38</td>
</tr>
<tr>
<td>The Respondents</td>
<td>39</td>
</tr>
<tr>
<td>The Responses</td>
<td>40</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>52</td>
</tr>
<tr>
<td>The Common Ground</td>
<td>52</td>
</tr>
<tr>
<td>The Payers</td>
<td>53</td>
</tr>
<tr>
<td>The Providers: Hospital Groups</td>
<td>55</td>
</tr>
<tr>
<td>The Providers: Professional Groups</td>
<td>61</td>
</tr>
<tr>
<td>Other Stakeholders</td>
<td>66</td>
</tr>
<tr>
<td>THE FUTURE OF VBP</td>
<td>68</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>71</td>
</tr>
<tr>
<td>Appendix 1: Glossary</td>
<td>74</td>
</tr>
<tr>
<td>Appendix 2: P4P literature</td>
<td>76</td>
</tr>
<tr>
<td>Appendix 3: Stakeholder Questions</td>
<td>77</td>
</tr>
</tbody>
</table>
BACKGROUND

Introduction

For years, policymakers have searched for ways to control spiraling costs to the public health care sector while improving the level of quality provided. The reason for the focus on inpatient services is clear: the average hospital inpatient admission costs between $2000 and $3000 a day.\(^1\) The reimbursement structure inadvertently allowed for reimbursement of duplicate tests and services as well as for care resulting from poor performance such as readmissions after avoidable adverse outcomes.\(^2\) Medicare began as a fee-for-service (FFS) program, reimbursing providers based on each service performed and therefore obviating any incentive to curtail medical expenditures for procedures such as cardiac catheterizations and diagnostic tools such as CT scans and MRIs.

While expenditures have been increasing, quality of care still lags behind that of other countries. In 2000, the Institute of Medicine (IOM) published *To Err is Human: Building a Safer Health System*, arguing that the US health care system did not take enough measures to insure patient safety and recommending urgent measures to address this problem.\(^3\) The IOM’s 2001 report, *Crossing the Quality Chasm: A New Health System for the 21st Century*, boldly argued that the US health care system did not meet established benchmarks for quality.\(^4\) These reports, along with embarrassing anecdotal stories of poor quality and multiple studies and essays describing the problems with health care in the US, have led to nationwide efforts to improve quality. However, while the scope of information about, and research pertaining to, health care quality in the US has increased dramatically, serious questions concerning the quality of health care in the United States remain.
The Medicare Modernization Act (MMA) of 2003 and Reporting Hospital Quality Data for Annual Payment Update (RHQDAPU)

In 2003, Congress passed the Medicare Modernization Act, which President George W. Bush claimed would "allow the biggest improvements in senior health care in nearly 40 years.” Most of the public and media attention was on the massive $400 billion prescription drug benefit that accompanied the bill. However, Section 501(b) of the MMA also called for a new hospital performance-reporting program called the Reporting Hospital Quality Data for Annual Payment Update (RHQDAPU), a pay-for-reporting (P4R) program that would provide incentive for hospitals reimbursed by the acute Inpatient Prospective Payment System (IPPS) to report performance data on all hospital patients (not just Medicare patients). The goal of RHQDAPU was to increase transparency of health care performance and give health care consumers certain tools to make informed decisions about their health care choices. Beginning in 2004, hospitals would provide data on 10 different quality indicators of care for acute MI, heart failure, and pneumonia.

In 2005, the Centers for Medicare and Medicaid Services (CMS) launched the Hospital Compare website to allow anyone to view a comparison of these quality indicators across all participating hospitals. Although participation in RHQDAPU is technically voluntary, a revision to IPPS would allow CMS to reduce IPPS payments by 0.4% to hospitals that did not participate. This ensured that the vast majority of hospitals would submit performance data for public analysis and hospital comparisons. In 2005, almost every hospital (98.3%) did in fact release these quality data, and 96% met the required criteria and received full reimbursement. This initiative ran the risk of providing too little data for the typical consumer to use and
punishing hospitals for several arbitrary benchmarks. Certain indicators, especially the antibiotic requirement within 4 hours of arrival for pneumonia, also led to gaming (to be discussed later).

**Deficit Reduction Act (DRA) of 2005**

Through the Deficit Reduction Act of 2005, Section 5001(b), P.L. 109-171, Congress authorized the Department of Health and Human Services to develop a plan to implement Value-Based Purchasing (VBP), a policy mechanism that would link payment to performance, helping changing Medicare from a passive payer into an active purchaser. Congress asked CMS propose a plan by 2007 to implement VBP by 2009 and include 1) the continuous development of quality measures for the inpatient setting, 2) collection and validation of this data, 3) the structure for appropriate reimbursement (VBP adjustments), and 4) public reporting (transparency). Congress also directed the secretary of DHHS to adopt performance measures introduced by the 2005 report from the Institute of Medicine (IOM) entitled *Performance Measurement: Accelerating Improvement.* This report, following in the footsteps of *To Err is Human* and *Crossing the Quality Chasm,* argued that despite significant interest and widespread investments, as well as per capita health care spending far exceeding that of other countries, progress in improving health care quality remains slow, and significant disparities within our countries remain. The report criticizes our health care system for lacking a coherent, efficient system for assessing and reporting performance data as well as strong central leadership, without which individual stakeholders cannot act together to effect health care system reform. In order to create a comprehensive and standardized system of reporting performance data, IOM encourages the use of 22 Hospital Quality Alliance (HQA) measures as well as the Hospital Consumer Assessment of Health care Providers (HCAHPS) patient survey (described later). In response,
CMS began to investigate methods for creating a VBP payment program for inpatient reimbursement based on P4P.

The DRA also increased CMS's ability to ensure participation in RHQDAPU. Again, while participation in RHQDAPU was technically voluntary, a revision to the IPPS would allow CMS to reduce IPPS payments by an even greater margin (2%) for hospitals that did not participate. Meanwhile, CMS expanded RHQDAPU from 10 to 21 quality indicators regarding care for acute MI, heart failure, and pneumonia, as well as the Surgical Care Improvement Project (SCIP), which involves the use of prophylactic antibiotics for surgery.

The 2007 update to RHQDAPU, and the HCAHPS survey

CMS provided a 3.3% increase in Medicare funding for all hospitals that complied with RHQDAPU. For those that did not comply, the reimbursement increase for 2008 was only 1.3%. CMS also revised the hospital outpatient prospective payment system (OPPS) to add several additional measures to RHQDAPU, bringing the total number of quality indicators measured by RHQDAPU to 27. Measures include mortality measures, additions to SCIP, and results from the HCAHPS patient survey (as encouraged by the IOM Performance Measurement report). This survey, consisting of 27 items, introduced the concept of patient satisfaction as a benchmark, giving the patients the opportunity to influence Medicare reimbursement. HCAHPS includes satisfaction with communication with health care providers, responsiveness of staff, cleanliness, pain control, and discharge information.
The Premier Hospital Quality Incentive Demonstration (HQID) Project

In 2003, CMS initiated the Premier Hospital Quality Incentive Demonstration (HQID), a three-year pilot program designed to test the ability of P4P to improve the quality and decrease the cost of inpatient hospital care. To date, HQID is the largest pilot demonstration of P4P in the US. The demonstration was designed to "define quality measures and reward top-performing hospitals" and enrolled over 260 hospitals on a voluntary basis. The demonstration tracked performance data on 34 quality measures in five different clinical categories and developed by numerous public and private organizations: acute myocardial infarction (MI), heart failure, community-acquired pneumonia, coronary artery bypass graft (CABG), and hip/knee surgery. Among the measures are shorter length of stays and fewer readmissions. CMS would then compare hospital performance data with the best recommended practices for all 5 categories and reward or punish hospitals according to a carrot-or-stick model. For example, hospitals in the top 10% of each category would receive an additional 2% reimbursement per case, while hospitals in the next 10% would receive an additional 1% reimbursement per case. However, hospitals failing to improve from the bottom 20% by the third year face a 2% reduction in payment. For certain hospitals with slim margins, a gain or loss of 2% in payment can be significant.

CMS’s proposed plan to implement inpatient P4P for Medicare: Value-Based Purchasing

In November of 2007, CMS placed before Congress a 104 page report outlining guidelines for the adoption of its proposed VBP initiatives, built around the concept of P4P and covering all inpatient services reimbursed by Medicare. The report proposes replacing RHQDAPU with a new program that combines public reporting and an incentive-driven payment
system that focuses on quality of health care, patient-centeredness, and efficiency, as encouraged by the IOM in its 2006 report, *Rewarding Provider Performance: Aligning Incentives in Medicare*. To address clinical quality as well as efficiency, and having learned from the mistakes of Premier HQID, CMS promised to expand outcome measures, while noting that risk adjustment would be more complicated than process measures and that CMS must be vigilant in finding negative unintended consequences of incentivizing outcomes. To address patient-centered care, CMS would incorporate the HCAHPS patient survey. Importantly, CMS proposed that Congress implement VBP in a budget-neutral fashion.

The new VBP plan would include several important components. CMS would develop a Performance Assessment Model to assess hospital performance across various quality indicators, which would then be compiled into a Total Performance Score. This Total Performance Score would then be translated into an incentive payment. Since few would expect such sweeping reform to be perfect in initial attempts, the measure development process should be well-structured and spelled out clearly. CMS must have a plan to validate the data and study the impact of the transition so that any problems can be quickly addressed. The transition itself should be phased over several years. And most importantly, all the results should be publicly available.

**The Performance Assessment Model and Measure Development/Selection**

The performance assessment model would be the yearly overall indicator from which incentives are determined. The performance assessment model would combine several domains such as clinical processes, clinical outcomes, and patient satisfaction. Policymakers could choose to weigh each domain equally, or instead choose to give greater weight to some domains,
depending on what they wanted to emphasize. CMS would set the attainment threshold, at which point hospitals would earn attainment points (and incentive payments), at roughly 50th percentile, with the benchmark for exemplary performance (and full incentive payments) set at roughly 90th percentile. CMS believed that the amount of payment incentive should not be determined by a linear exchange rate: for example, if policymakers wanted to emphasize improvement at the lower end of performance, the slope of the exchange rate would be higher at that point, thus giving greater incentive to lower-performing hospitals to improve performance. In order to prevent punishment of hospitals performing at a lower baseline level and provide incentive for all hospitals to improve, the proposed model would use the higher score between an "attainment score" (the traditional performance score) and an "improvement score" (which compares the hospital with its prior baseline performance).

Importantly, not all measures would have to be used to determine incentive. While all VBP measures would be reported publicly, only a subset would determine incentive payments. CMS could use the other measures for measure development and selection. In this fashion, CMS could assess the suitability of various measures without having potentially adverse effects in terms of incentive payments. The process of developing and selecting measures would occur in three stages. In the first stage, a proposed measure would undergo a "preliminary data submission period" in order to collect data. In the second stage, CMS would release the data publicly and investigate whether the measure met a list of criteria (such as feasibility, importance, and scientific acceptability). In the third stage, CMS would add the measure to the VBP criteria for incentive payments. Using the same process, CMS could retire certain measures deemed inappropriate.
Phased transition from RHQDAPU to VBP

CMS prefers a phased transition from RHQDAPU to VBP in order to give hospitals time to adjust to measures and benchmarks, as well as to gather enough performance data to determine how to set incentive payments. CMS would require three years to make the transition, incorporating a “baseline” period and a “measurement” period. CMS, responding to hospital feedback, noted that the benchmarks and thresholds should be based on prior experience rather than current experience (as per Premier HQID), mainly so that the hospitals would know the performance standards in advance. The baseline period, starting on April 1 and ending March 30, would precede the measurement period by 12 months, allowing CMS to set baseline performance scores for all the hospitals as reference points. The measurement period would then precede the VBP year (from October 1 to August 30) by several months, allowing all hospitals to submit data (and resubmit if necessary), as well as giving CMS the time to obtain and validate the measurements and then determine incentive payments. CMS could introduce new measures experimentally in this fashion, allowing hospitals to collect data without public reporting. While RHQDAPU allowed 135 days for hospitals to submit data after each quarter, which made it difficult for CMS to adjust certain measures and provide feedback, VBP would narrow the submission period to 60 days and possibly require reporting for every month.

Funding of incentive payments in a budget-neutral fashion

A key component of VBP would be budget neutrality. Like Premier’s HQID program, CMS would require no additional funding for this model. Instead, CMS proposed that the incentive come from a pool created by a percentage of the existing DRG payment and linked to performance based on clinical services provided during an inpatient stay. In other words, if
hospitals performed well, they would receive their percentage back, plus an additional amount. On the other hand, if hospitals performed poorly, they might not receive their percentage. CMS noted that the size of incentive would be important in dictating hospital behavior, though no definitive research has been conducted to determine the optimal size. CMS proposed a 2-5% of DRG payments be used as incentive payments. CMS also proposed that the incentive payments apply to all discharges, not only the discharges with specific incentive measures, so that hospitals would work to improve their quality of care across the board and not just for a subset of patients. One important example of this would be the HCAHPS patient satisfaction survey. Given that not all hospitals would earn the full incentive payment, CMS proposed that the unused funds be placed in a pool to be dispersed later as additional incentive. Not surprisingly, hospitals have voiced displeasure with budget neutrality, given that gains made by some hospitals would be at the expense of others (to be expanded later).

**Supporting research and hospital feedback**

To create and provide an evidence base for the VBP model, CMS created a Hospital VBP Workgroup that investigated measures, data infrastructure and validation, incentive structure, and public reporting. To assist in this research, this workgroup contracted with RAND, an independent nonprofit global policy think tank, as well as a team from Brandeis University. The Workgroup reviewed P4P literature and surveyed existing P4P programs. Upon finding that P4P literature was scarce and limited, the Workgroup had RAND interview a range of organizations including P4P sponsors, organizations involved with public reporting, hospitals participating in RHQDAPU and Premier HQID (the large Medicare inpatient P4P demonstration), and hospital associations.
According to RAND, all hospitals reported that RHQDAPU and Premier HQID were helpful in preparation for VBP, which most believe to be inevitable. Most hospitals wanted the measures used to be evidence-based and with the support of quality organizations such as the National Quality Forum (NQF) and the Hospital Quality Alliance (HQA), and hospitals tended to be reticent about allowing patient satisfaction data to help form their bottom line. RAND also noted that RHQDAPU caused hospitals to anticipate the coming emphasis on quality of health care. However, hospitals tended to complain about the cost of implementation of health IT required for these programs, and they questioned whether the incentives provided could even cover the costs, much less earn them bonuses. Hospitals especially dislike relative thresholds for performance measures and almost universally prefer absolute thresholds, such that every hospital reaching and exceeding this benchmark would garner a reward without any hospital necessarily having to “lose.” Small and rural hospitals have voiced concern that smaller volume of patients and the cost of health IT would make participation more difficult, and that VBP should include measures that focused specifically on these hospitals, such as transfers and coordination of care.

My hypotheses are that payer groups and provider groups have significant differences in their approach to the VBP proposal, especially concerning issues such as budget neutrality and the process of measure selection and development. Nevertheless, I believe that all parties believe VBP is inevitable and that Congress will ultimately pass the VBP proposal sooner rather than later.
LITERATURE REVIEW

Search Strategy

It is important to understand the evidence base that does – or does not – underlie CMS' assumptions about what P4P and VBP can accomplish. I assessed the evidence by performing a PubMed search with the key terms (pay for performance) and quality, with limits set to full text and free full text. The search returned 68 results, of which all 68 had abstracts. I then reviewed abstracts for appropriateness of inclusion; for example, I excluded studies performed outside the United States, and I also excluded opinion, editorial, and "perspective" pieces, or other content not related to the present study question. I included actual trials of quality improvement interventions, and I included abstracts that were sufficiently unclear that I needed to review the full article. This process produced a list of six articles including actual trials of improvement in the United States relevant to this paper's research question. I also searched the CMS VBP Proposal to Congress for relevant papers and selected the Premier White Paper, given that it is a preferred citation of P4P proponents, and Grossbart et al., given that it used contemporary controls for comparison. The final systematic literature review includes 9 studies, and they are summarized in Appendix 2.

Premier HQID studies

Of this review's eight studies, four directly involve the CMS Premier HQID program. The studies involving Premier HQID are of most interest to me because the structure of Premier HQID is similar to that of the VBP proposal. Given the importance of Premier HQID and its implications on the design of VBP, it is surprising that investigators have conducted so few
studies on the effectiveness of Premier HQID. In addition, the data overall suggests weak improvement in quality among participating hospitals, and even then, the studies themselves have flaws that bring the question the internal and external validity of these results.

The Premier White Papers are reports summarizing Premier HQID’s year 1 and year 2 performance data for all 262 participating hospitals. Instead of using contemporary controls, Premier elected to use a before-and-after design, comparing the quality performance data at the beginning of year 1 to the same measures at the end of year 1 for the entire universe of initial participants. For the second report, Premier merely compared the mean scores for year 1 with the mean scores for year 2. Participants were all invited and chose to accept the invitation; CME did not elect to collect data on similar measures in nonparticipating medical centers, most likely because such data are not always collected with incentives or requirements to do so. Premier collected data on performance measures for five different conditions (AMI, HF, CAP, CABG, and hip/knee replacement), including both process measures and inpatient mortality data. Measurement of the performance measures appears to be equal, valid, and reliable, decreasing the possibility of measurement bias.

The Premier White Paper found that from the beginning of year 1 to the end of year 1, participating hospitals on average improved in all categories, from 3.4% in AMI to 9.8% in CAP. All differences were significant. Between year 1 and year 2, the mean performance improved in all categories, from 3.5% in AMI to 9.8% in CAP. As with year 1, all differences were significant.

The internal validity for the Premier White Paper is poor to fair. All participating systems used the same measures, with well-established and agreed-upon underlying meanings.
However, the Premier systems were not compared to nonparticipating systems. Without contemporary controls, we cannot know whether the improvements in the Premier demonstrations were the result of factors other than incentive payment. For example, the improvement could be due to the public nature of the program (transparency) or some technical aspect of the program (health IT and consultation support, for example). Nonetheless, the fact that more than 200 participating systems showed improvement over the period is suggestive.

The external validity for the Premier White paper is unknown. The hospitals participating in Premier HQID were invited and accepted the offer, demonstrating particular motivation to participate in a P4P that is likely not shared by all hospitals across the nation. These hospitals were also likely better prepared from a structural standpoint than the average hospital. Therefore, it is difficult to support the use of these data when trying to justify the creation of VBP that is based on similar principles and design.

Lindenauer et al. conducted a non-randomized controlled trial studying the ability of participation in the Premier HQID program to improve quality of care for AMI, HF, and CAP.15 Lindenauer et al. gathered data on 207 hospitals participating in the Premier HQID and compared them to a control group of 406 hospitals under public reporting only (through the Hospital Compare website) over 2 years. Hospitals wishing to participate in Premier HQID needed a minimum of 30 cases per condition annually to be eligible. Of the 266 hospitals participating in Premier HQID, 11 dropped out during the first 2 years, leaving 255 hospitals eligible for primary analysis. For controls, Lindenauer et al. screened for eligible participants among the HQA voluntary reporting program. Controls were eligible if they submitted at least 30 cases for each condition annually and did had not either declined participation in HQID or
participated in HQID and dropped out later. Thus, crossover was not an issue. Lindenauer et al. attempted to match each HQID participant with one or two HQA hospitals based on number of beds, teaching status, region, location (urban versus rural), and ownership status (profit or non-profit).

Based on Table 2, the intervention and control groups appear comparable, with very little variation in characteristics such as number of beds, location, teaching status, region, and ownership. However, the participants were not selected at random, instead being matched by the aforementioned characteristics. There was also no mention of allocation concealment, so I assume the investigators were aware of the allocation process. Given that the selection process was not randomized, there is potential for selection bias, but the groups are comparable per Table 2, suggesting that selection bias is less of a factor.

Primary outcomes were the differences between 10 HQID measures in the 4th quarter of 2003 and the 3rd quarter of 2005. The measures included 5 AMI measures, 2 HF measures, and 3 CAP measures, as well as appropriate care measures for all three conditions and composite process scores for all three conditions plus an overall composite process score for all 10 conditions. Assessments of these measures appear to be equal, valid, and reliable, with little possibility of measurement bias.

While randomization is difficult to impossible in a study such as this, the lack of randomization nevertheless raises the possibility of confounding. The investigators did not describe the process of selecting the group of 207 hospitals in the intervention group, except to say that they were chosen because they matched closely with control hospitals by demographics. The investigators did mention that they excluded controls that had either turned down HQID participation or began participating and later dropped out, thus controlling for any confounding
based on the possibility that these hospitals felt like HQID was not a workable plan or a good idea. However, the hospitals that have agreed to participate in Premier HQID are likely not identical to the control groups. Given that they were invited and voluntarily decided to participate, these hospitals are more likely to have the infrastructure sufficiently developed to support P4P, as well as staff and employees more likely to embrace P4P. The authors attempted to account for this participation effect by adding hospitals that declined participation in HQID. However, these hospitals are likely also not comparable to the control, based on the same reasons provided by the authors above, as well as the fact that these hospitals were not incentivized and therefore would not have experienced the same effects of P4P.

Lindenauer et al. found that both P4P hospitals and control hospitals showed improvement over the 2 year study period, but also that P4P hospitals showed greater improvement on 7 of 10 individual measures, ranging from .6% (p=.009) to 10.9 (p<.001). In addition, P4P hospitals showed improvement on composite measures, from 4.1% (p<.001) to 5.2% (p<.001). Also, stratified analysis demonstrated an inverse relationship between baseline performance and overall improvement in both groups. After adjusting for effects of baseline performance, the overall improvement decreased to 2.6% for AMI (p<.001) and 4.1% for HF (p<.001).

Overall, internal validity of this study is fair but not good, mainly because it is difficult to believe that the hospitals in the intervention group are identical to the hospitals in the control. One way to control for the so-called “volunteer effect” – that is, the constellation of variables that make one hospital willing to try a challenging new program, while another similar hospital will not – may have been random assignment of groups, but this would have been impossible in this type of study. I am not convinced that analysis including the hospitals that declined to
participate would be enough to account for this effect. Another issue is that the investigators did not control for technical aspects of Premier (health IT and consultation). The results themselves are statistically significant as demonstrated by the p-values. Lindenauer et al.’s study is to date the largest Premier HQID study with a contemporary control (including a large majority of Premier HQID programs), which does help internal validity.

External validity is also fair. The hospitals agreeing to participate in both the Premier HQID and the HQA voluntary reporting program may be a select population of hospitals motivated to be the pioneers of quality control. Indeed, their difference from controls should be taken as a reason for further exploration of what makes institutions willing to innovate. These results may be hard to extend to the general population of hospitals that may be less well-prepared to handle either voluntary reporting or P4P, unless we can identify the "willingness to innovate" variables and then find ways to spread them to other institutions. Also, the improvements, while statistically significant, are small and may not justify the large amount of financial investment into the program. Importantly of note is that this study received funding from CMS. I am unsure what, if any, role CMS had to play in the study other than financial assistance.

Grossbart et al. conducted a non-randomized controlled trial studying the ability of participation in Premier HQID to improve quality of care for AMI, HF, and CAP. Grossbart et al. compared 4 acute care hospitals in Catholic Healthcare Partners participating in Premier HQID with 6 acute care hospitals in the same organization that declined to participate in Premier HQID. To ensure comparability, Grossbart et al. chose hospitals in this group that had similar levels of service. All 10 hospitals have an open-heart program and are referral centers. In
addition, Grossbart et al. tried to choose hospitals with similar average annual discharges, case mix index, and staffed beds. Grossbart et al. did not choose the hospitals randomly.

Per Table 1, the intervention and control group appear somewhat comparable, but the small number of hospitals in each group mean a very wide standard deviation in each category (staffed beds, annual inpatient admissions, and case mix index). The averages of each characteristic seem somewhat similar. Also, given that all 10 hospitals are from the same group, it is possible that the demographics of the hospitals are similar, but Grossbart et al. provided little data to support this theory. Given that each of the four hospitals in the intervention group come from different regional health systems, the demographics for these hospitals likely vary.

Grossbart et al. did not specify how they selected the hospitals for the control group other than to say that they were chosen for similar characteristics as the intervention group.

Since the selection process was not randomized, there is fair to strong potential for selection bias. The groups do not seem entirely comparable per Table 1, and the Grossbart et al. did not provide many details on demographics and patient admissions or patient comparability. Compounding the problem is that the group sizes are very small (4 and 6 hospitals). Thus, the potential for selection bias is very strong. Primary outcomes included composite quality scores on AMI, CAP, and HF, as well as an overall composite quality score. The authors then calculated improvement between Oct 1, 2002 and Sept 30, 2003. These measurements appear to be equal, valid, and reliable, diminishing the possibility of measurement bias.

The potential for confounding is high, given the lack of randomization, though like the other studies, randomization would be difficult to impossible to achieve. In particular, the selection process raises the potential for confounding. Grossbart et al. report that the four hospitals in the intervention group volunteered, while the others that were included in the control
group declined to participate. Even more so than Lindenauer et al., this raises the strong possibility that the priorities of these hospitals for quality improvement are different, given that they all come from the same hospital group. The hospital administrators likely consulted each other before making their decisions, suggesting that participation (or declining to participate) in each of these groups was not an independent event. This suggests a systematic difference in the level of preparation for, and commitment to, quality improvement, both in the infrastructure and the hospital staff.

Grossbart et al. found significant differences in improvement from baseline to one year after intervention between the P4P hospitals and the control hospitals. For HF, the P4P hospitals improved 19.2% compared with 10.9% for control hospitals (p<.001). However, for AMI and CAP, P4P hospitals improved by 2.9% versus 3.1% and 7.9% versus 7.2% (p not given, but reported as not significant). Overall, due to the strength of the improvement for HF, P4P hospitals improved by 9.3% versus 6.7% (p<.001).

Given the small sample size and strong possibilities of selection bias and confounding, the internal validity of this study is poor. Grossbart et al. made little attempt to control for confounding, and beyond several superficial demographics, also made little attempt to prove that the hospitals in each group were comparable to one another. Since the study only included 4 hospitals in the intervention group and 6 in the control group, the power of this study is very low as well. In addition, as in the previous studies mentioned, Grossbart et al. did not control for the contribution of technical aspects of Premier HQID to the improvement. Given that this study involved only hospitals from a specific hospital group with specific goals of quality improvement, external validity of this study is also poor.
Glickman et al. conducted a non-randomized controlled trial investigating the effects of participation in Premier HQID on quality of care regarding AMI only. Glickman et al. compared patients from hospitals participating in CRUSADE (a voluntary observational quality-improvement initiative) and Premier HQID to patients from hospitals participating in CRUSADE only. Overall, 54 hospitals were in the intervention group, compared with 446 hospitals as the control. Glickman et al. included patients if they presented at a CRUSADE hospital within 24 hours of ischemic symptoms plus positive cardiac biomarkers or ischemic ST-segment ECG changes.

Per Table 1, the two groups are comparable in the types of patients presenting as well as patient history and presenting signs and symptoms, but the facilities are not as comparable. The P4P hospitals are larger, more academically oriented, and located primarily in the south. The study design was obviously not randomized, given that the investigators used all the Premier HQID participants also participating in CRUSADE as their intervention group, and using all other hospitals in CRUSADE as the control. Given that the selection process was not randomized, there is potential for selection bias, but Glickman et al. made no attempt to control for these variables.

Glickman et al. used a combination of 6 CMS process measures and 8 non-CMS process measures, as well as in-hospital mortality. These measures appear to be equal, valid, and reliable, thus minimizing the possibility of measurement bias. Like the other papers, lack of randomization introduces the possibility of confounding. Given that the demographics of the hospitals are not very similar, the hospital groups are not too comparable. Surprisingly, the patient variables were similar. Nevertheless, the potential for confounding is fairly high and not explored in detail.
Glickman et al. found that differences in process measures as well as patient in-hospital mortality were not statistically significant, with the only exceptions being aspirin at discharge, smoking cessation counseling, and lipid-lowering agent at discharge (3 of 15 measures). Like Lindenauer, stratified analyses found that hospitals with lower baseline rates improved faster, but there was no statistically significant difference in improvement between P4P hospitals and control hospitals.

Glickman et al.’s study has fair internal validity due to the lack of comparability between hospitals in the intervention versus control group, introducing the possibility of both selection bias and confounding for which Glickman et al. did not attempt to control. Glickman et al.’s study, like Lindenauer et al.’s study, suffers from the “volunteer effect,” given that the hospitals in the intervention group voluntarily enrolled in Premier HQID, suggesting a difference in priorities and infrastructure regarding quality improvement. Unlike Lindenauer et al., however, Glickman et al. does not make any attempt to control for the volunteer effect. Glickman et al., just like the other studies, did not control for the contribution of technical aspects of the program to quality improvement. Glickman et al.’s study has fair external validity as well. This study, like those conducted by Grossbart et al. and Lindenauer et al., used hospitals that voluntarily enrolled in Premier HQID, limiting the study’s applicability to hospitals across the nation.

These four papers demonstrate how mixed the data supporting Premier HQID really is. While the Premier White Papers purport statistically significant improvements for all five conditions through year 1 and year 2, the lack of contemporary control makes concluding that P4P led to quality improvement to be very difficult. Grossbart et al. found statistically significant differences for all three conditions studied, but a small size along with problems
involved with the selection process and little attempt to control for confounding variables leads to poor internal validity. On the other hand, Glickman et al. found no statistically significant differences in performance between the hospitals in Premier HQID and CRUSADE and hospitals in CRUSADE only. Glickman et al.’s study, however, suffers from similar issues of selection bias and confounding, though a much larger sample size adds some power and increases internal validity. Lindenauer et al. found statistically significant differences in quality performance. Of the three papers, Lindenauer et al. appears to have the strongest internal and external validity. However, while Lindenauer et al. does the best job of controlling for confounders, the study cannot account for all possibilities of confounding. All four papers rely on voluntary participation of hospitals in Premier HQID, and the selection process is not random, decreasing internal and external validity. Finally, none of the papers controls for possible contributions of the technical aspects supplied in Premier HQID. Premier HQID provided health IT for data collection as well as considerable technical support for participating hospitals, which could very well have made a statistically significant different in performance. Clearly investigators must continue to research the ability of Premier HQID to improve performance while also doing a better job controlling for confounding.

All the papers about Premier HQID, except for the Premier White Papers, investigate the difference between quality improvement produced by P4P and public reporting as compared to public reporting only. Many experts believe that public reporting provides strong non-financial incentive to improve quality. However, none of these papers attempts to provide control hospitals that are neither involved in Premier HQID nor publicly reporting. This would have allowed investigators to determine just how much of a contribution public reporting was making to quality improvement as opposed to some other confounding variable. In addition, the public
reporting for the P4P groups is not identical the public reporting for the control; the intervention groups report to two different programs instead of just one.

Not surprisingly, Lindenauer et al. and Glickman et al. observed an inverse relationship between baseline performance and overall improvement throughout the study. This is an interesting finding, given that Premier HQID only rewards attainment and not improvement (as the VBP proposal would). It is possible that the VBP’s joint attainment/improvement scoring system may lead to even faster improvement among hospitals with lower baseline performance, but this is only conjecture.

Other P4P papers

I also examined other studies more generally related to P4P. In these studies, P4P's effectiveness is similarly mixed, and the programs may not have been of the same (or even similar) design to VBP, indicating that their results might not be as applicable to future VBP considerations. Nevertheless, I was still interested in reading papers that studied the ability of P4P to improve quality of health care in principle.

Cutler et al. performed a retrospective cohort study on the ability of P4P in a California-based medical group, Mercy Medical Group, to improve the quality of cholesterol management on an outpatient basis. Cutler et al. identified 165 patients with diabetes mellitus type 2 and elevated LDL-C cared for under a P4P program (CDCM) and compared them to a control group of 1694 patients in the same plan with the same conditions, but under routine care. Based on Table 4, the two groups appear roughly comparable, though the study only listed gender and age for comparison, so selection bias is still an issue. Confounding is a possibility since no eligibility
criteria were given for patients under CDCM program. Cutler et al. measured rates of LDL-C testing and goal attainment of LDL-C (<130mg/dL) between the two groups. These measures appear equal, valid, and reliable, so measurement bias is unlikely. The study showed that participants in CDCM were at >75th percentile for LDL-C testing rate and >90th percentile for goal attainment of LDL-C, compared with <20th percentile and 75th percentile for routine care. Both differences were statistically significant (p<.001).

Internal validity for this study is fair to good. This author would have liked more demographic information for comparability as well as criteria for participating in CDCM, as this could lead to possible confounding. External validity is limited since this study involved simple outpatient measures, and VBP is a large inpatient P4P program.

Mandel et al. studied the effect of P4P on 44 practices in Cincinnati as part of an asthma improvement collaborative. Mandel et al. did not use a contemporary control for this study. Mandel et al. measured flu shot percentage, controller medication percentage for children with persistent asthma, and written self-management plan percentage and compared performance over the network from the beginning to the end of the 3 year study period. Mandel et al. monitored the distribution of rewards earned by 43 practices based on their performance on the aforementioned measures. The study found that the percentage of asthma patients receiving “perfect” care rose from 4% to 88%, while the percentage of asthma patients receiving the flu vaccine rose from 22% to 62%.

Internal validity of this study is fair. On one hand, this paper lacks a contemporary control group. However, the need for a control group in this type of study is probably diminished by the fact that any contemporary control groups would likely not improve at the
same rate, if at all, given a lack of obvious incentive (such as a public reporting system for performance results, as is the case in other studies such as Lindenauer et al.). As with the Cutler et al. study, external validity is limited for VBP since this study involved outpatient measures rather than inpatient measures. In addition, this study only emphasized positive rewards, whereas VBP is a budget-neutral design.

Rosenthal et al. was a longitudinal study of the ability of P4P mechanisms to influence the quality of care involving outpatient measures in California physician groups. Rosenthal et al. used data from PacifiCare Health Systems' P4P programs in California to determine quality of care in 10 clinical and service quality targets involving cervical cancer screening, mammography, and HbA1c over one year. As a control group, Rosenthal et al. used data from PacifiCare Health Systems' groups in the Pacific Northwest (Oregon and Washington). Rosenthal et al. did not provide comparison tables so it is difficult to determine the comparability of these different groups. Rosenthal et al., however, did note that there were "no statistically significant differences in the trends between the 2 networks before the quality incentive program." Rosenthal et al. found that, of the three categories, only cervical cancer screening resulted in a statistically significant difference between the two groups (absolute difference 3.6%, p=.02). Given that PacifiCare paid $3.4 million in bonuses, Rosenthal et al. concluded that this P4P program was not an efficient use of resources.

Internal validity is fair. The two groups are difficult to compare given that Rosenthal et al. provided no demographic data concerning either group, though Rosenthal did state that pre-program treatment trends were similar. Regarding VBP, external validity is fair, given that VBP is an inpatient program with a budget neutral design.
Doran et al. examined the effects of outpatient P4P programs on family practices in the UK.\textsuperscript{22} Doran et al. used data from 8105 family practices in England in the first year of the P4P program to measure the performance for quality indicators in 10 clinical domains, including asthma, COPD, hypertension, and diabetes. Doran et al. was also interested in the relationship between exception reporting (the means of excluding patients from P4P calculations) and overall performance. Doran found that median achievement (the proportion of eligible patients for whom targets were achieved) was 83.4 percent, ranging from 90.1% for DM to 96% for hypothyroidism. Exception reporting was low (median 6%) but ranged from 0% to 85.8%, with 1.1% of practices excluding more than 15% of patients. Doran et al. believed that median achievement was high because the benchmarks of attainment were too low. Doran et al. also believed that gaming the system was uncommon, but widespread among a handful of practices (given the exception reporting data).

Internal validity of this study is fair. Without any control groups, it is hard to isolate the amount of quality improvement is due to the P4P incentives. Regarding VBP, external validity is fair to poor, given that this is a different type of program (outpatient, positive incentives) in a different health care system.
Overview of the Senate Finance Committee Roundtable on Hospital Value-Based Purchasing

On March 6, 2008, the Senate Finance Committee, led by Senators Max Baucus and Charles Grassley, co-sponsors of the Medicare Value Purchasing Act of 2005, convened a roundtable of participants representing a variety of health care viewpoints and with strong vested interest in the outcome of VBP. Senator Grassley stated that “this trajectory of spending” could not continue, and that VBP was one way to “bend this long-term growth” while “ensuring that the patients get the best possible care.” The Senate Finance Committee was looking for diversity in opinions and suggestions regarding VBP, stating that in striving “for a system that provides health care that’s safe, effective, patient-centered, timely, efficient, and equitable,” much remained “to be discussed, explored, and decided.”

Representatives of academia (Dr. Meredith Rosenthal of the Harvard School of Public Health), patient groups (the Commonwealth Fund), quality endorsement (National Quality Forum [NQF]), payer groups (CMS, Blue Cross Blue Shield of America [BCBSA], National Business Group on Health), hospital groups (Association of American Medical Colleges [AAMC], American Hospital Association [AHA], the Billings Clinic, Federation of American Hospitals [FAH], Alegent Health, and Premier Health Alliance), and professional groups (American College of Surgeons [ACS], the American Nursing Association [ANA], the American Organization of Nurse Executives [AONE], and the Society of Hospital Medicine [SHM]), industry groups (the American Hospital Association, the Blue Cross Blue Shield Association, the Federation of American Hospitals, Alegent Health, the National Business Group on Health, and
Premier Inc.) were all present. Interestingly, the AMA was not present at the Roundtable, and we raise possible reasons for their exclusion elsewhere in this paper. The Senate Finance Committee asked specific questions involving performance standards, quality measures, structure of incentives, and implementation. The participants used this valuable opportunity to voice their concerns about and suggestions for CMS's VBP proposal to the Senate Finance Committee. Some participants chose to answer each question one by one, whereas others used their time as a platform for protecting and promoting the interests of their particular constituents that might be threatened or not well-represented by the new plan.

Performance Standards

Only four respondents even mentioned minimum thresholds, with only one (Premier) in favor. Two respondents (Dr. Rosenthal and BCBSA) discouraged the use of minimum thresholds, as minimum thresholds would “discourage the poorest performing hospitals from engaging in quality improvement.” Two respondents (Dr. Rosenthal and BCBSA) discouraged the use of minimum thresholds, as minimum thresholds would “discourage the poorest performing hospitals from engaging in quality improvement.”

Five respondents endorsed CMS’s proposed combination of attainment scores and improvement scores. ACS emphasized the importance of improvement over baseline attainment, concerned that “high performance” rewards would exacerbate the physician shortage. Dr. Rosenthal also suggested an alternative to CMS’s proposal, stating that prorating a fixed reward by percentage of achievement would eliminate arbitrary cutoffs for reward.

Quality Measures

The vast majority of respondents trusted in the NQF in the process of measures selection and development. Of the 14 respondents addressing this issue, 12 respondents named the NQF
in response to the question about the process used to develop, test, refine, endorse, adopt, and retire quality measures. Three respondents (NQF, ANA, and BCBSA) only mentioned the NQF specifically as the organization of choice to be involved with measures selection and development, while six respondents identified the mechanism currently in use by the CMS for its reporting program (measures endorsed by the NQF, approved by the HQA) as the preferred mechanism for VBP. Three groups (Commonwealth Fund, Billings, and Dr. Rosenthal) identified the Joint Commission for possible involvement, while three groups (National Business Group, Commonwealth Fund, and Dr. Rosenthal) named the NCQA as part of the process. In addition, two groups (Commonwealth Fund and Dr. Rosenthal) mentioned the AHRQ as a possibility as well. One group, the ACS, believed that the process used by the NQF to endorse measures was “onerous and too slow” and needed to “mature to meet the needs of a true value-based purchasing system.”

As for the types of measures used, only three respondents, Dr. Rosenthal, Billings, and the Commonwealth Fund, specifically emphasized the importance of process measures, though the relative silence on the question of process measures could have resulted from their extant emphasis in P4P programs such as the Premier HQID. Four respondents mentioned outcomes measures, with the Commonwealth Fund noting that outcomes were “the bottom line in health care.” However, both Dr. Rosenthal and the Commonwealth Fund urged caution with the use of outcomes in VBP, given that outcomes measures could “make the VBP system vulnerable to inappropriate incentives” such as adverse selection, while Premier urged the exclusion of 30-day mortality measures because they “do not currently provide adequate feedback for hospitals to evaluate their performance.” Two respondents (the Commonwealth Fund and Dr. Rosenthal) also mentioned the inclusion of structural measures. Six respondents stressed the
importance of patient-centered measures such as HCAHPS, though Premier urged caution with HCAHPS, indicating that "more experience needs to be obtained."29 One respondent, the ACS, asked for measures that were longitudinal and measured efficiency. Two respondents (ACS and FAH) endorsed the use of composite measures in order to be "direct drivers of outcomes,"27 while Dr. Rosenthal warned that all-or-none measures make sense "only if there is a clinical reason to believe that there is little value to each measure at the margin, unless they are all accomplished together."26 In response to the question of how often CMS should review measures, three groups (AAMC, the Commonwealth Fund, and Dr. Rosenthal) stated annually, while the BCBSA stated 18-24 months, based on its own program for Blue Distinction Centers.

Structure of Incentives

Of the seven respondents who addressed the initial size of the incentive, only the Commonwealth Fund believed that CMS’s proposed 2-5% of base DRG payment was reasonable. Four respondents believed that the reward should be 2% or less; the AHA claimed that 2-5% would be “excessive for an untested system,”30 while others cited the effectiveness of smaller rewards in other CMS programs such as Hospital Compare and Premier HQID.29 Four respondents stated that incentives should be applied to specific conditions as opposed to all DRGs, while one respondent, the Commonwealth Fund, argued that certain broader measures that were not condition-specific could be better applied at the provider level.31

Eight respondents indicated the importance of aligning hospital and physician incentives, but they offered few suggestions as to how to accomplish this. Dr. Rosenthal indicated that "there is no published evidence on this question" but suggested that "paying attending physicians for improving the same evidence-based processes or reducing inpatient complications would
encourage productive collaboration of a kind that is unlikely to occur otherwise." The FAH suggested that the NQF was investigating the problem, whereas the NQF stated that coordination of measure development should assure that measures “roll up” (e.g., infection rates calculated the same way), be setting-neutral, apply common conventions, and aggregate in composites or summary metrics.

Eight respondents directly or indirectly opposed the principle of budget neutrality in VBP. None of the respondents endorsed the concept of budget neutrality. BCBSA argued that budget neutrality would “make it difficult for low performers to improve in succeeding years.” The AHA stated that “any perception that these efforts are about budget cutting and not performance improvement will sour the kind of change that everyone would like to see in care delivery.” The AONE warned that any VBP system “used punitively and arbitrarily to reduce payments to hospitals” would “be devastating to hospitals and have negative consequences... to the nursing staffs... and the communities they serve.” The FAH believed that “the overall incentive payment should ensure that sufficient funds be targeted at the lower performing hospitals.” Finally, according to Premier, VBP should “create a positive incentive to improve performance.”

Implementation

All seven respondents addressing the issue of included hospital types asserted that CMS must make special consideration for rural and safety-net hospitals. Two respondents (FAH and Premier) insisted that CMS must establish some special arrangement involving minimum number of cases, with the FAH arguing that CMS should allow opt-outs for conditions with fewer than 25 cases. Four respondents believed that the CMS should provide some sort of
assistance for rural and safety-net hospitals in the form of education, technical support, or financial assistance.

Where They Sit: Advocacy

14 respondents seized the opportunity to advocate for their constituents’ interests concerning VBP. The AAMC emphasized that teaching hospitals were special institutions whose situations VBP might not recognize adequately and in fact potentially harm. On 28 occasions, the AAMC extolled the benefits of teaching hospitals to the health care system, particularly emphasizing their role in research and caring for complex and severely ill patients, as well as the difficulty in measuring performance adequately for these special cases. The AAMC stated that “teaching hospitals are unique and vital to health care delivery and scientific discovery” and “disproportionately treat complex and severely ill patients who often present with multiple co-morbidities and serious, or rare, complications of routine medical conditions.” In addition, the AAMC stated that they “provide unique services not found at other hospitals including transplantation, trauma and burn care, participation in clinical trials, and other services” and also “serve as a safety net” and “train future physicians.” The AAMC feared that “VBP poses unique challenges,” especially concerning the “uniform approach to Hospital Acquired Conditions” and the fact that “POA coding may be neglected because it sacrifices life-saving care.” The AAMC also worried about how VBP would address the complexity of care and teaching arrangements in teaching hospitals. The AAMC concluded by hoping that “VBP does no harm to the nation’s teaching hospitals.”

The AONE used the roundtable as a platform to address the severe shortage of nurses in the US on 24 different occasions. The AONE cited 13 studies and reports indicating the harms
of nursing shortages as well as the benefits of an enlarged nursing workforce.\(^{34}\) The AONE emphasized that “nursing is a key partner in sustaining quality outcomes” and warned against treating nursing as a fixed cost, as the current system of DRGs does. The AONE concluded by stating that nursing and nursing care have “taken on critical importance” in efforts “to ensure patient safety at every level of care.” The ANA promoted nursing interests, urging for the inclusion of nursing sensitive measures on 12 different occasions. The ANA stressed that VBP “fails to include the central provider (the nurse) and fails to recognize nurses' extensive contributions” and insisted that “implementation of VBP without reporting nursing indicators would be incomplete.”\(^{37}\) Meanwhile, SHM reminded the Committee of the prominent role that hospitalists had in quality improvement. On 11 occasions, the SHM mentioned that hospitalists “serve key leadership roles” and lead quality improvement projects concerning best practices, discharge planning protocols, and quality measure development.\(^{38}\)

Both the ACS and AHA worried about the unintended consequences of an unproven plan for their constituents. The ACS, on 15 different occasions, emphasized that VBP was “still in early stages of development,” current data did not support the efficacy of this new system, and that CMS had inadequately developed the currently proposed system.\(^{27}\) The AHA, on 7 occasions, feared that the proposal contained too many unknowns, stating that “the report lays out a variety of options but stops short of recommending a specific design,” that “pay-for-performance is a policy that is still largely untested and unproven,” and urged Congress to “move forward cautiously, mindful of unintended consequences.”\(^{30}\)

Both the Billings Clinic and the Commonwealth Fund took advantage of the opportunity to promote the interests of rural and underserved populations. The Billings Clinic, on 7 occasions, described rural providers as eager to participate and brought major positive attributes
with their integrated models of care, but that lower volumes would place a premium on process over outcomes. The Commonwealth, on two occasions, promoted the needs of vulnerable populations with their concern that a reliance on outcomes would make VBP vulnerable to inappropriate incentives, such as avoidance of sicker patients.

Finally, both NQF and Premier used this opportunity to advocate for their continued inclusion on future CMS programs. The NQF, on six different occasions, wished to protect its prominent role in endorsing measures for VBP, stating that “NQF endorsement... has become the ‘gold standard’ for health care performance measures” and that “major health care purchasers, including CMS, rely on NQF-endorsed measures to ensure that measures are scientifically sound and meaningful and to help standardize performance measures across the industry.” Premier, which held a pivotal role in data collection and management in the Premier HQID program, defended its potential role in VBP on three occasions by stating that “part of the success of the HQID project ... can be attributed to the support provided for project participants” and that “Premier has facilitated the distribution of best practices amongst all participants and provided consulting support for hospitals with lower levels of performance.”

Where They Stand: Support for VBP

The lone representative of academia, Dr. Rosenthal, supported VBP. Dr. Rosenthal believed that the “proposed hospital incentive program represents another step toward a more enlightened role for Medicare as an agent for delivery system improvement” and that “the alternative (allowing payment and performance to remain unrelated) is no longer tenable.” Dr. Rosenthal felt that VBP “should proceed based on sound principles and evidence as to what works.” The NQF also supported VBP. NQF “strongly” urged Congress to pass VBP
legislation, believing that there existed "much potential to enhance the quality of health care," and "standardized performance measurement and reporting system is a fundamental building block for creating a national health care system that provides high quality service and is affordable and accessible to all Americans." The lone representative of patient interests, The Commonwealth Fund supported VBP as well. The Commonwealth Fund stated that there was "reason to believe that substantial progress in both quality and efficiency should be possible" and that the "the quality of care across areas does not correspond to spending." The Commonwealth Fund added that "VBP should reflect a broad view of value" and that "paying for high performance rewards achievement."

The payer groups generally supported VBP. BCBSA was "concerned that health care is unaffordable" and supported "changing incentives to promote better care." BCBSA noted that "providers are generally paid based on services provided" and asserted that "incentives must be changed." BCBSA concluded that "implementing VBP would help change reimbursement from paying for more to paying for quality." The National Business Group, the most enthusiastic supporter of VBP among the industry groups, "strongly (urged) Congress to pass legislation that would implement pay-for-performance on a widespread basis in the Medicare program," believing that "it is necessary for the financial future of Medicare as well as for the quality and safety of care... that pay-for-performance be used to harness the government's leverage as the largest purchaser of health care in the US to move Medicare and all other payers towards paying for effective health care and quality outcomes rather than units or volume of services." The group emphasized that "too often payment made without regard to whether services are needed" and hoped that "CMS is taking the next step towards moving from being primarily a passive
payer for health care to an **active purchaser** for health care, using its enormous power to buy the best possible care for millions of beneficiaries. *Emphasis in original*]

The hospital groups were much less supportive of VBP. The AAMC was concerned that VBP was not ready for full adoption, stating that the use of existing coding systems for complex patients could have “serious, unintended consequences” and that “such a program will present challenges both to hospitals and CMS to implement.” The AAMC added that “there remains considerable need for progress on a research agenda for studying financial incentives in health care.”

The AHA expressed concern about the untested nature of VBP and unintended consequences that might result from widespread implementation. The AHA stated that “the report lays out a variety of options but stops short of recommending a specific design” and urged Congress “to move forward cautiously, mindful of unintended consequences.” The AHA added that “past initiatives were expected to resolve hospital quality issues, but more work remains to be done.” The AHA asserted that “the additional implementation of a value-based program will be challenging and resource intensive.” The AHA expressed concern that “CMS Acting Administrator Kerry Weems indicated that Congress could achieve budget savings by retaining some or all of the unspent reward funding, which would mean that implementation of the pay-for-performance plan could create a ‘backdoor’ budget cut for hospitals” and warned that “any perception that these efforts are about budget cutting will sour change.” The AHA conceded that “using incentives is an important concept worth exploring” but insisted that “pay-for-performance is a policy idea that is still largely untested and unproven.”
Billings also was reluctant to endorse VBP. Billings noted that VBP “may well present the opportunity to improve quality and reduce costs,” but cautioned that VBP “will require significant refinement” and “a plan for policy makers to stick with a Value-Based Purchasing framework ... over 10-15 years will take vision and persistence, something often difficult to achieve in a complex political environment.” The FAH expressed similar reluctance to endorse VBP. The FAH considered itself a “strong proponent of quality and performance measurement.” The FAH “supports moving to a payment system that links payment to performance” but urged Congress “to proceed carefully.” The FAH felt that “our national quality infrastructure does not have sufficient methodologies for determining if patient outcomes are better because of process measures being employed as part of patient care routines.”

Of the hospital groups, only Premier supported VBP. Premier was “highly concerned that VBP is intended to generate $1.65 billion in savings to the Medicare program over five years,” believing instead that “VBP must be used to drive change and create incentives for continuous quality improvement” through “quality gains... reinvested in hospitals participating in the program.” Nevertheless, Premier believed that “Congress should act now to enact a hospital value-based purchasing program” because “accelerating health care performance improvement is imperative to patients and our economy.”

A majority of professional groups did not support VBP. The ACS, in particular, was a very strong critic of VBP. The ACS warned that VBP was “still in early stages of development,” with “few data indicating measures are moving us toward our goals.” The ACS also complained that “current measures sets still require more testing,” and VBP measure sets “currently lack measures that are longitudinal.” In addition, the ACS felt that “patient-shared
decision-making is a critical missing link.” Also, the ACS believed that the amount of incentive “has been adequately addressed.”

The ANA and AONE also took cautious stands. The ANA made no explicit mention of support, though it did insist that CMS “re-examine payment structure and recognize the value of nursing” by including nursing sensitive measures. The AONE asked that VBP “be fair and improve quality and performance,” and also emphasized that VBP “not be used punitively and arbitrarily to reduce payments to hospitals.” The AONE, as a subsidiary, aligned its message somewhat closely with that of the AHA, though the AONE did not include any comments about the lack of readiness of VBP.

SHM was the only professional group in support of VBP. The SHM, an enthusiastic supporter of VBP, stated that it “shares (CMS’s) commitment to improving the efficiency and the quality of our health care system.” The SHM agreed with CMS in that “reforms that align Medicare payments more closely with the quality of care provided, rather than reward volume of services as the current system does, are urgently needed” and added that “America’s hospitalists stand ready to work with you in achieving this important objective.” In addition, SHM “supports CMS’s efforts to implement policies designed to promote the delivery of care that is safe, effective, timely patient centered, efficient, and equitable” and concluded that “SHM strongly supports CMS’s proposal to reward hospitals based upon performance improvement and attainment of a minimal quality threshold.”
STAKEHOLDER RESPONSES

Rationale

While I found the responses to the VBP Roundtable to be very useful given that they represented a wide range of health care interests and generally answered the same set of questions, I also decided to draw responses directly from key stakeholders in the health care industry. I wanted the opportunity to interview these stakeholders, none of whom had previously seen the interview questions, in order to obtain spontaneous and non-scripted responses. In addition, I hoped to obtain viewpoints from these stakeholders that the VBP Roundtable questions did not address.

Methods

I created a list of questions about P4P and VBP and submitted the questions for IRB approval, at which point the IRB declared his project exempt from IRB approval as it was not human subjects research. I identified candidate respondents who were key stakeholders involved with or influenced by the VBP program, and who represented a variety of positions and sides of the issue. I then submitted a standardized letter via e-mail to each candidate respondent asking for his or her permission for an interview about CMS’s inpatient P4P initiatives and the VBP proposal. Upon receiving a positive response, I arranged a mutually agreeable interview time with the respondent’s assistant over the phone or in person. Before the start of the interview, I informed each respondent of my intent to record the interview using a digital voice recorder, transcribe the interview, and use the respondent’s name, title, and direct quotes from the interview for the paper, and asked for consent from the respondent for each of these conditions.
Once I received consent, I began the interview (see the interview protocol in Appendix 3). The interviews ranged from 22 minutes to 90 minutes in length. Upon completion of the interview, I offered to send the respondent a copy of the transcript and a copy of the completed paper. I searched through the responses for each question and coded the responses relative to each prompt, measuring the respondent’s relative stand on each issue, using occurrences of key phrases and ideas in the response. For example, I might code responses based on approval or disapproval of a certain concept raised in the prompt, such as budget neutrality. Afterwards, I used a third-party reader not involved with this paper to review a small sample of the coding results independently. Our intercoder agreement was 95%. Where there were disagreements, we identified the sources of disagreement and resolved the differences in each case.

The Respondents

The respondents are Jeffrey Spade, Vice President of the North Carolina Hospital Association; Nancy Foster, Vice President for Quality and Patient Safety Policy of the American Hospital Association; Brian Goldstein, Executive Associate Dean for Clinical Affairs, UNC School of Medicine, and Chief of Staff, UNC Health Care System; and Thomas Valuck, Medical Officer and Senior Adviser in the Center for Medicare Management of the CMS. Mr. Spade speaks from the viewpoint of hospitals, especially rural and underserved hospitals. Ms. Foster also represents the viewpoint of hospitals, though in a more nationwide and global fashion. Dr. Goldstein speaks as an academic and a representative of physicians. Finally, Dr. Valuck represents the payer perspective, particularly the perspective of the largest payer, the Centers for Medicare & Medicaid Services.
The Responses

Three of the four respondents (Mr. Spade, Dr. Goldstein, and Dr. Valuck) felt that Medicare was a logical place in which to implement a pay-for-performance payment system. Mr. Spade commented that “Medicare has been one of the innovators in terms of payment structure for health care” and therefore was an appropriate setting “to introduce innovation to payment structure and reimbursement system.” Dr. Goldstein felt that Medicare was “a reasonable place to start.” Dr. Valuck said that CMS, as “the biggest payer,” had “a fiduciary responsibility for the Medicare trust fund” and was “seeking to provide leadership in VBP.” Ms. Foster, however, while agreeing “that Medicare is, as the largest payer of hospital services, is in many respects a logical place to do VBP,” warned that “significant challenges” within a “very complicated payment system that has more pages of regulation than the IRS” as well as payments that fail to “get you back up to your total cost” in addition to any penalties “may mean you aren’t able to deliver care that needs to be delivered.”

When asked to identify supporters of P4P, two of four respondents (Mr. Spade and Ms. Foster) identified payer groups as likely supporters. Mr. Spade said that “the payers themselves would be one group because they want to be sure they’re getting the highest value,” while Ms. Foster stated that “the main advocates thus far have been certainly been employers and other purchasers other than Medicare who believe that essentially you get what you pay for.” Two of four respondents (Mr. Spade and Dr. Goldstein) identified providers and physicians as most likely to be opposed to P4P on principle. Mr. Spade stated that “physicians don’t like the constraint of performance system that requires that they be judged by standards” since it “seems constraining to patient care.” Dr. Goldstein believed that “physicians in small practices, especially rural practices, and physicians more generally may object depending on the
documentation, data collection, (and) recording.” On the other hand, Dr. Valuck disagreed with the concept of advocates and opponents in reference to P4P, stating instead that “we all share the goal of using correctly-aligned incentives to encourage better quality and avoid unnecessary costs in care.”

Three respondents (Mr. Spade, Ms. Foster, and Dr. Valuck) took neutral stances on the question of physician buy-in. Mr. Spade stated that “physicians can see the value in a payment system that optimizes performance” but added that “they’re skeptical of whether the government can do it right and make it work well.” Ms. Foster believed that physician understanding of P4P was “for the most part limited at this point” and that “physicians across the continuum really need to understand what’s going on and its implications to patient care.” Dr. Valuck believed that physicians “agree with the premise that payment should reward better quality of care and their efforts to avoid unnecessary costs of care” but added that physicians “question whether we have the right measures and the right infrastructure yet for them to be successful under that kind of program.” Dr. Valuck did say that VBP measures “are all measures that have been developed by professionals for professional accountability.”

In response to whether P4P had the ability to reduce cost while improving quality, three of the four respondents (Mr. Spade, Ms. Foster, and Dr. Valuck) agreed. Mr. Spade believed P4P had “great potential to improve quality,” citing Premier HQID as evidence that P4P “can drive down costs as well as improve aspects of outcome performance like mortality (and) complications like mistakes,” though he cautioned that “I would never claim that everyone has it figured out right.” Ms. Foster agreed that P4P “has that potential,” believing that “it has a number of opportunities to really incentivize people to provide care that is not only high quality but also eliminates unnecessary steps in health care” and “drive down costs.” Ms. Foster also
issued the caveat that "it's not universally true that whatever one would want to incentivize
would necessarily drive down costs, at least not in the short run." Dr. Valuck also agreed that
"VBP can increase the quality of care while avoiding unnecessary costs," adding that P4P held
"a lot of potential in beneficial effects, including... rewarding those who are successful in
enhancing efficiency in care, and also providing data for public reporting, which informs
consumers and also creates a bit of a competitive incentive to perform well vis-à-vis your peers."
Dr. Goldstein, while believing that "some of the measures that are included in most P4P
programs can achieve that (goal)," such as "preventing a hospital-acquired infection," also
presented a counterexample of "managing diabetics," which might require "more medication,
more office visits, (and) more frequent phone calls."

All four respondents agreed that P4P programs had the potential to lead to unintended
consequences such as adverse patient selection and data gaming. Of the four respondents, only
Ms. Foster allowed that unintended consequences were "very much a concern." Ms. Foster
stated that "comments... were made (by the AHA) last year when CMS was floating its VBP
white paper" stating that "regardless of how you define the system, there's a potential for
unintended consequences." She believed that "a responsible approach required a strong and
ongoing evaluation of the system and its impact in order to prevent unintended consequences."
Mr. Spade agreed that there was "always the potential" for unintended consequences "when you
adjust the system to create a certain outcome," adding that a system like P4P might encourage
"fraudulent" behavior. However, Mr. Spade added that "I think (fraudulent behavior is) at the
margin" given that "Medicare and Medicaid have a history of understanding what's appropriate
patient care and how to audit that and a good structure in place to do that." For example,
according to Mr. Spade, "DRG was subject to gaming for a long time... (and now) there are
audit structures in place to watch for fraudulent behavior.” Mr. Spade also added that “adverse selection is minimal,” given that “physicians are controlling the admissions processes” and “it would be very difficult for a physician to say, ‘I’m not admitting this difficult patient to this hospital.’” Dr. Goldstein stated that while “the potential is there,” he added that “it’s not clear to what degree hospitals have the ability to manipulate the system that way even if they want to.” Dr. Goldstein concluded that the potential for unintended consequences had “a lot has to do with the way the program is structured.” Dr. Valuck, while agreeing that “any payment policy has the potential for unintended consequences, especially with a relatively new application of a concept,” also asserted that P4P “has a lot of promise for actually addressing some of the current unintended consequences of our payment system, particularly misaligned incentives.” Dr. Valuck agreed with Dr. Goldstein that the potential for unintended consequences “depends on how VBP is applied and how it’s monitored.” Both Mr. Foster and Dr. Valuck agreed that a potential means of avoiding unintended consequences would be to align hospital and physician behavior. Mr. Foster said that “hospital P4P needs to be aligned with physician behavior that’s being driven by a physician payment structure,” while Dr. Valuck stated that “aligning measures across various settings of care would lead to more harmonized incentives and better care coordination.”

None of the respondents believed that EMRs were required for the successful implementation of P4P programs, but three respondents (Mr. Spade, Dr. Goldstein, and Dr. Valuck) believed that EMRs would certainly be helpful. Mr. Spade believed that while he “(did not think) it’s always necessary to achieve good outcomes,” he also had “no doubt that the more information you have that allows you to design great optimal care for patients, and that includes health IT.” Dr. Goldstein stated that EMRs were “not required… but they help with case
finding, they help with data collection, they help with... putting data together.” Dr. Valuck agreed, stating that EMRs were “not required” but “would certainly facilitate the adoption and effectiveness of VBP models.” Ms. Foster, on the other hand, felt that there were “enormous issues around trying to embed (EMR) in a VBP system,” specifying that “not all hospitals can afford IT” and that “history around implementation of EMR in hospitals” suggested that there were “far too many places where they brought a system up and the physicians found it cumbersome and refused to use it.”

In response to whether data transparency was important for success of P4P programs, Dr. Valuck stated that “publicly accessible information is highly desirable” because “public reporting provides a non-financial incentive for health care providers to improve their performance.” In addition, Dr. Valuck said that transparency “provides very useful information for decision-making, for consumers and for purchasers to make better informed decisions about health care that’s being provided in their communities.” Dr. Valuck added that transparency would help create “a competitive playing field” through “non-financial incentives” that “encourage providers to focus on improving their performance.” On the other hand, both Mr. Spade and Dr. Goldstein believed that transparency was helpful but not necessary. Mr. Spade stated that transparency was “necessary to drive improvement” but “isn’t necessary to gain improvement,” since “improvement will happen without transparency because people will be driven by the payment system.” However, Mr. Spade added that “transparency can drive greater value out of improvement.” Dr. Goldstein felt that Hospital Compare was “not bad because it lists every individual measure, and most of those are process measures,” which in turn told the viewer “a little about the kind of care that’s being provided.” However, Dr. Goldstein expressed great concern about “all-or-none scoring (or) optimal care scoring,” which to Dr. Goldstein “is a
bad thing and takes it too far.” Dr. Goldstein believed that getting “the same amount of credit or blame for each individual measure” did not “accurately (reflect) the quality of care.”

Three of four respondents (Mr. Spade, Dr. Goldstein, and Dr. Valuck) agreed on the importance of patient perception of care. Mr. Spade stated that “patient satisfaction is key to performance and has real value” but added that “I’m not willing to say we have it right” and suggested that patient satisfaction measures might “have to be tweaked quite a bit.” Dr. Goldstein felt that patient satisfaction was “a legitimate variable” and that “the patient’s loyalty determined by how they perceive the way they were treated should be part of the equation.” Dr. Valuck felt that patient experience data was “very important” and that performance assessment should address “all of the six key dimensions of quality.” However, Ms. Foster, while believing that patient satisfaction was important, felt that HCAHPS was so new that its use could lead to “unintended consequences” and that CMS would “need experience with the measure to understand if it is measuring what you think it’s going to measure.” Ms. Foster added that HCAHPS surveys were not offered in languages other than English and Spanish, thus excluding “sections of care on a whole important part of the patient population.” Ms. Foster also used the idea that HCAHPS might be “rewarding (hospitals) for new construction” of buildings as an example of “things that you want to think about when you design a P4P system to reward hospitals.”

Three respondents (Mr. Spade, Ms. Foster, and Dr. Valuck) believed that small and rural hospitals faced special disadvantages from VBP. Dr. Goldstein declined to answer as he believed himself not to be qualified. Mr. Spade argued that rural hospitals required government investment in the form of health IT and “helping hospitals invest in systems of care.” All three respondents agreed that small volume was a critical problem for small and rural hospitals. Mr.
Spade stated that with small volume, “your variation can be broad.” Ms. Foster said that “it’s really problematic with small sample size to know what a rewardable performance is.” Dr. Valuck added that “when numbers of cases are too small, the results can be skewed based on small numbers related to particular instances may not be generalizable in their populations.” All three respondents also agreed that CMS might need to adjust the P4P structure in order to accommodate small and rural hospitals. Mr. Spade stated that “there needs to be more thought how P4P works for those hospitals” and that “it’s up to the government to understand how P4P affects those organizations differently, so they can meet the challenges of P4P systems.” Ms. Foster said that “the measures aren’t well-designed to measure the care largely provided in small and rural hospitals” and “it does require more thought than just applying the same system in the same way that you would to a 250-bed community hospital.” Dr. Valuck suggested that, as one option, CMS “could select measures specifically relevant to small hospitals, look at a longer time frame, let’s say two years of data instead of one, to presumably get twice as much data,” or consider “various statistical techniques.”

When asked about the goals that the Premier HQID program meant to accomplish, Mr. Spade felt that they included “the concept of hospitals working together, sharing information and implementing great design principles for patient care.” Ms. Foster believed that Premier HQID “was intended to look at whether hospitals would respond to incentives in payment and accelerate their improvement in performance as a result.” Dr. Goldstein thought that “the goals were to demonstrate that care could be improved without increasing costs.” Dr. Valuck stated that the primary goal was “to test whether financial incentives and public recognition – the financial and non-financial incentives – could be effective in stimulating quality improvement.” Overall, all four respondents were very enthusiastic about the accomplishments of Premier
HQID. Mr. Spade noted that through the Premier HQID program, “there was less variation, overall increase in performance in terms of outcome, and a general improvement in cost.” Mr. Spade added that “by working together and creating some formal structure to share information,” hospitals “learn from each other and make changes in their performance,” though Mr. Spade cautioned that “I’m not sure that the performance incentive is structured correctly to drive the investments necessary to drive performance increase” and that the hospitals “figured out how to work together to drive performance, and I’m not sure the payment structure is what drove it.”

Ms. Foster believed that Premier HQID “absolutely marvelous job showing that a group of highly motivated group of hospitals intent on quality can succeed.” Dr. Goldstein believed that Premier HQID “made progress demonstrating the fact that you can improve care without substantially increasing cost of care.” Dr. Valuck, while noting that the results were still preliminary, noted that “the demo showed pretty dramatic positive results, even against a background of general improvement likely stimulated by the advent of hospital public reporting on Hospital Compare.”

All three respondents asked about measures (Mr. Spade, Ms. Foster, and Dr. Goldstein) agreed that VBP should include process measures. Mr. Spade stated that “there’s no doubt process measures make sense.” Ms. Foster agreed that “the best measures to hold hospitals accountable include process of care measures.” Dr. Goldstein noted that “process measures have gotten refined enough at this point” that they were “hard to argue with.” Mr. Spade, Ms. Foster, and Dr. Goldstein all agreed that VBP should also include patient satisfaction measures. Mr. Spade stated that “if we don’t start including (patient satisfaction measures), we’ll never learn what aspect of that information should be driving performance.” Ms. Foster agreed with including “patients’ experience of care... where clearly it’s within (the hospitals’) control.”
Goldstein felt that patient satisfaction measures were “reasonable to be part of the equation” as long as they were similar to the HCAHPS survey. All three respondents were hesitant to endorse outcomes measures as part of VBP. Mr. Spade believed that investigators “need to figure out more about outcomes” and felt unsure that “mortality is the only outcome we require.” Ms. Foster questioned whether outcomes measures were “actually showing and whether or not one would be rewarding based on something truly within the control of the hospital.” Ms. Foster also worried that “it could easily lead to unintended consequences... where people are less eager to admit patients who are several ill or likely to have complications of care.” Dr. Goldstein believed that outcomes measures would experience problems with “risk adjustments and the kinds of patients you accept in the first place.”

Two respondents (Mr. Spade and Ms. Foster) recommended that NQF endorse all measures for VBP. Mr. Spade appreciated the “quasi-public” structure of NQF and thought that “the structure to (endorse measures) well is in place.” Ms. Foster also approved the use of HQA as a supplier of measures for public reporting. In response to incentivizing attainment versus improvement, Mr. Spade believed that incentivizing improvement was “a good idea in transition,” but after a transition period, the government would “want a system that clearly rewards high performance.” Dr. Goldstein thought the combination of attainment and improvement scores was “a big plus” and “makes it a lot easier for hospitals to collaborate.”

Two of three respondents (Mr. Spade and Dr. Valuck) believe that there is enough evidence to support VBP’s effectiveness. Mr. Spade cited the Premier HQID program as evidence to the ability of P4P to improve process and decrease variation and indicated that “we’re at that natural evolution point where we need to start moving in that direction (of P4P).” Dr. Valuck stated that there was “so much intuitive appeal and so much good early evidence
from CMS' demos and private sector experience that CMS is proceeding forward.” Dr. Valuck, however, also added that CMS needed to “perform evaluation and monitoring to make sure that we don’t somehow dilute the positive aspects with potential unintended consequences.” While Dr. Goldstein “(didn’t) think we can wait double-blind placebo-controlled trials for everything,” he also stated that “the new system has to be nimble enough to adjust quickly.” Dr. Goldstein illustrated this by arguing that if data showed that the practice supported in VBP’s measures was “not the ideal treatment,” and a physician instead used the latest data for a different treatment, “you’ll get dinged on that measure.”

On the important issue of budget neutrality, two respondents (Dr. Goldstein and Dr. Valuck) believed that VBP could be viable as a budget neutral program, while the other two respondents (Mr. Spade and Ms. Foster) did not. Dr. Goldstein believed budget neutrality was “reasonable” and an “incentive for all of us to try to work at high quality but also efficiently.” Dr. Goldstein did admit that he would “rather see this direction than some other worse alternative, which would be rationing of some other kind.” Dr. Valuck defended budget neutrality by citing economic literature that stated that “the most effective incentive occurs when an individual or organization has lost something that they need to earn back,” which he stated provided “more powerful incentives than simply adding a bit more to what the individual or entity receives that they may not have counted on anyway.” Mr. Spade disagreed, stating that the government must “make investments now to drive savings in the future.” He added that “if we do not recognize that that investment is necessary, then we will not accomplish the improvement that we are seeking” and that if “the reimbursement system just moves payments from low performers to high performers, I think that that’s an illogical approach.” Ms. Foster agreed, stating that VBP “should be a system that provides rewards for hospitals” and that VBP
would be “not survivable if in fact it is about taking money away and cost cutting to hospitals.” Like Mr. Spade, Ms. Foster felt that savings through “reduction in complications in care, the reductions in readmissions, the reductions on ongoing health care needs in those patients” would result in “newfound savings to actually reward improved performance.”

Two of four respondents (Mr. Spade and Ms. Foster) believed that budget neutrality could have a harmful effect on hospital collaboration. Mr. Spade believed that budget neutrality “forces the health care providers to compete with each other in a way that’s counterproductive” and did not believe that “that drives the type of improvement we want.” Mr. Spade added that “anything that helps hospitals work together is better than systems that drive us apart” and that “budget neutrality is a way to drive providers apart.” Ms. Foster cited quality literature as evidence that “nothing has been as effective as these collaborations in driving quality improvement forward to high levels of attainment with a great deal of speed,” and she worried that increased competition would lead to a “chilling effect on hospitals’ willingness to share their secrets to success with each other.” The other two respondents, Dr. Goldstein and Dr. Valuck, did not believe that budget neutrality, in practice, has harmed hospital collaboration. Dr. Goldstein had “seen mainly a lot of collaboration between hospitals” and concluded that “there’s mostly positives right now” surrounding hospital collaboration in a budget neutral payment program. Dr. Valuck noted that hospitals participating in the Premier HQID program “readily shared best practices” and that any possible harmful effect on collaboration “hasn’t been our experience.”

Only one of three respondents, Dr. Goldstein, believed that the size of the incentive proposed by CMS (2-5% of DRG payments) was “a good range” and “enough to get the hospitals’ attention.” Ms. Foster believed that 2-5% “is a little high to start with” for “an
untested program” where literature has demonstrated “a pretty mixed result, as to whether P4P programs have actually attained their established goals.” Ms. Foster instead advocated for 1%, while doing “continuous ongoing assessment on the impact that P4P is having.” Mr. Spade felt that CMS “(did not) have enough experience” though he would only say that “clearly there’s only a small percent necessary to drive participation.”

When asked about the future of VBP, all three respondents (Ms. Foster, Dr. Goldstein, and Dr. Valuck) believed that VBP was inevitable. However, only Ms. Foster was willing to guess that “within the next two years we’ll see some movement” with “debates and decisions next year.” Both Dr. Goldstein and Dr. Valuck, on the other hand, thought that adoption of VBP, while inevitable, would not be fast. Dr. Goldstein believed VBP would experience a “slow adoption.” Dr. Valuck agreed, stating that VBP adoption would be “incremental but steady.”
DISCUSSION

The Common Ground

Despite differences on details, a majority of stakeholders agree on certain aspects of the VBP proposal. Most recognize that P4P has the ability to decrease costs while simultaneously increasing quality in many categories. Generally, every stakeholder at the VBP Roundtable and interviewed by this author believes that the implementation of P4P into Medicare’s payment system is inevitable and a logical idea at least in theory. All respondents believe that Premier’s HQID program was a success and a good model upon which to build VBP, and many respondents have noted the effect of collaboration between hospitals to improve when even a small incentive . Hence, most respondents support (or are at least resigned to) the implementation of P4P into Medicare’s payment system as a means of improving quality and efficiency in health care, and most respondents have been encouraged by at least several years of pilot testing through Premier HQID that P4P is a viable payment system to achieve these goals through VBP.

When asked about measures, virtually all respondents agree that process measures and patient centeredness measures are important for VBP, and most agree that CMS must perform more research regarding outcomes measures as these measures are vulnerable to unintended consequences such as adverse patient selection. In addition, CMS must carefully consider risk adjustment when dealing with outcomes measures. The vast majority of the stakeholders recognizes NQF as an appropriate endorser of measures for use by VBP, and most approve of the current method (measures endorsed by the NQF and selected by the HQA) of introducing measures for VBP.
Most endorse the incentive structure of a combined scoring system using both attainment and improvement scores as a fair method to reward a broad spectrum of providers. Premier HQID mainly rewarded those with high baseline performance and provided little incentive for hospitals with lower baseline performance to improve. However, the VBP proposal would also give those hospitals incentive by rewarding improvement as well. Respondents also feel that the use of minimum thresholds threatens to disincentivize providing superior quality of care. Most respondents believe that hospital and physician incentives must be aligned correctly for VBP to achieve its potential, but respondents offered few suggestions as to how to make achieve this goal, suggesting that the issue of alignment of hospital and physician incentives will be difficult to address.

When asked about implementation, virtually all stakeholders believe that VBP must include special considerations for small and rural/safety net hospitals, given their smaller patient pool and financial fragility. None of the respondents believe that either EMRs or data transparency are necessary for the success of VBP, but both would help improve the program. Most respondents have been impressed with the ability of the Hospital Compare website to elicit disclosure of data regarding performance on various quality measures.

The Payers

Despite a wide range of concordance among stakeholders on a variety of issues, differences in opinions between payer groups and provider groups are evident. Payer groups tend to be very interested in using VBP as a vehicle to reduce costs in addition to improving quality and efficiency of health care. Health care costs now consume roughly 15% of the nation's GDP and up to half of these costs may be due to unnecessary care. Most believe that
health care costs will only continue to rise until the economy can no longer sustain the costs. Given Medicare’s purchasing power, many in the industry are watching closely to see whether the VBP proposal comes to fruition. It should be no surprise that all payers welcome a change in payment system from the largest health care payer which they hope would help the push to change all other payer systems to P4P as well.

Not surprisingly, the payer groups are generally very supportive of CMS’s VBP proposal. Clearly, CMS, the sponsor of the VBP proposal, is in favor of transitioning Medicare from a passive payer to an active purchaser of care, and CMS believes that its VBP proposal will be restructuring the payment system in the right direction. CMS’s decision not to reimburse for so-called “never” events is a step in a similar direction of not reimbursing for unnecessary care or mistakes made during care. In formulating its VBP proposal, CMS wishes for “the right care for every person every time” and sees VBP as an opportunity to “improve quality and avoid unnecessary costs.” The idea of P4P, according to CMS, should be to decrease practice variation throughout the US and control skyrocketing health care costs. CMS has incorporated the lessons learned from Premier HQID in hopes of a better P4P program that CMS can effectively implement nationwide. VBP should no longer only benefit hospitals with high baseline performance, now that CMS has proposed a combined attainment/improvement score.

CMS also believes in the viability of a budget-neutral design, which has sparked tremendous criticism among provider groups. Certainly in an atmosphere of aversion to raising taxes, a budget-neutral design is more politically feasible. Dr. Valuck also believes that a budget-neutral design adds a stronger incentive for hospitals to improve quality compared to a design that only adds positive incentive. The Premier HQID program currently operates with a budget-neutral design, and CMS has used data published by Premier to support the claim that
P4P can, with a budget-neutral design, improve the quality and efficiency of health care. CMS will likely continue advocating for a budget-neutral design for VBP, as that should increase the chances that the Senate Finance Committee would find this design more palatable than other discussed reimbursement options.

The National Business Group on Health is the strongest supporter among the participating stakeholders at the VBP Roundtable. This group, which represents large private employers nationwide, “strongly (urges) Congress to pass legislation” involving VBP, given that Medicare often reimburses providers “without regard to whether services are needed.” Private employers have, just like CMS, watched health care costs increase steadily and would like VBP to “harness the government’s leverage as the largest purchaser of health care” in the US in order to “move Medicare and all other payers toward paying for effective health care and quality outcomes.” Payers are already well aware of the “amplifier effect” following CMS’s legally required policy no longer reimbursing for upcoding the DRGs for “never” events (in other words, gaining significant publicity leading to action). Private employers, believers in the ability of P4P to reduce costs, are clearly waiting for VBP to become law in order to benefit from Medicare’s “amplifier effect,” given that any action to change Medicare’s reimbursement system would have profound effects on the industry. Payers clearly see the opportunity to effect P4P programs as policy throughout private insurance, and they simply need for Medicare to exert its leverage in order to achieve widespread adoption.

The Providers: Hospital Groups

Hospital groups share similar goals with those of the payer groups with respect to increasing the quality and efficiency of health care while decreasing the ballooning costs
associated with Medicare. However, with respect to VBP, hospital groups differ from the payer
groups in their opinions concerning the readiness of the program for full-scale adoption in
Medicare, in addition to the structure of incentives and payments. Hospitals believe that CMS
should use VBP primarily as an engine for improvement in quality and efficiency of health care,
not as a method of cutting costs up front. Hospitals reason that a program well-designed to
improve quality and efficiency of health care should recognize savings in the long run that would
more than make up for initial investment.

Not surprisingly, in contrast to the payer groups, and as previously suspected, the hospital
groups are much more hesitant to support VBP than were the payer groups. Overall, the larger
organizations representing multiple hospitals, such as the AHA and AAMC, tend to be more
suspicious of VBP, while representatives of smaller hospital organizations, are less suspicious
(but still caution) of VBP. The greatest exception to this seems to be Premier, an organization
representing hundreds of hospitals and health care groups nationwide, though their interest in
VBP will be discussed later in the context of the resources created by the Premier demonstration
project.

The providers, including both hospitals and professionals, share the same goals of quality
improvement and improved efficiency as do the purchasers. Almost everyone agrees that overall
health care (and Medicare) costs have been spiraling out of control. In order to address this,
purchasers are in favor of plans that improve the quality provided through standardization of
care, while also increasing savings in the long run, since better quality of care is also cheaper
care. Improved quality of care would decreased expensive hospitalization times and decrease
morbidity and mortality as well, which would all result in decreased calls. At least in theory,
P4P would be able to accomplish these goals.
Hospital groups, however, have serious reservations about the design of VBP. One of the most contentious issues is the proposed budget-neutral design. CMS believes that VBP requires no additional funding to fund incentives, instead drawing incentives from a pool of an at-risk percentage of DRG payments. Dr. Valuck of CMS has cited evidence from economics literature stating that, in fact, hospitals should perform better if they believe they have “lost something that they need to earn back” as opposed to working toward incentives “that the individual or entity may have thought they weren’t going to get anyway.” Many provider groups have expressed fears that CMS, by advocating for budget neutrality, is trying to use VBP as a vehicle to cut costs rather than focusing primarily on quality improvement. Hospital and hospital group representatives argue instead that VBP will need an initial investment in order to drive the improvements in quality and efficiency that CMS desires. Hospitals also argue that a budget-neutral VBP program would punish lower-performing hospitals. Hospitals struggling to remain fiscally solvent may find even a 1% reduction in DRG payments to be backbreaking. In addition, hospitals lacking the funds

In addition, hospital groups fear that budget neutrality could potentially jeopardize helpful collaboration between hospitals. As Ms. Foster states, hospitals are worried whether increased competition would have a “chilling effect” on hospital collaboration, given that hospitals might be less likely to share secrets when they know they are competing for the same money. While Premier HQID has not shown such harmful competitive behavior, the amount of incentive (1% of DRG payments at risk) has also been relatively small. Now that CMS has proposed a larger incentive for VBP (2-5% of DRG payments at risk), hospitals may be even less inclined to collaborate on quality improvement efforts.
Hospital groups also fear that P4P has not shown the same sort of promise to improve quality that its proponents believe to exist. Indeed, this paper's mini-systematic review indicates that the evidence for P4P's effectiveness in respect to the Premier HQID program is mixed, and that the existing studies have design limitations. Certainly, hospital groups have been vocal about the relative lack of research that exists for P4P. The AHA cautioned that P4P is "a policy idea that is still largely untested and unproven." The AAMC, quoting Desai et al., notes that "the empirical foundations of pay for performance in health care are rather weak" and that there remains "considerable need for progress on a research agenda for studying financial incentives in health care." The hospitals clearly do not want a wholesale change in payment system in all aspects of health care until CMS has research that indicates that P4P has obvious benefits in health care quality improvement.

Similarly, most hospital groups also object to the untested nature of VBP. Both large and small hospital groups have expressed concern over the untested nature of VBP all expressing their hope that the government will implement VBP slowly and carefully. As discussed in the results, the FAH and Billings join the AAMC and AHA in emphasizing that VBP is "in the very early phase" and must be carefully developed in order to "avoid unintended consequences." These hospital groups clearly fear the implementation of a fairly unproven and untested payment system, given that the hospitals would likely bear the brunt of any problems experienced in a wide scale P4P adoption by Medicare. Not surprisingly, most hospital groups would prefer a slow adoption of a VBP program that will have plenty of field-testing and research performed and can adapt quickly to address administrative problems and rapidly evolving medical literature.
Even CMS’s VBP proposal admits that “there is little formal evaluation occurring of existing hospitals P4P programs” and that “most of the published studies have significant methodological limitations that hamper our ability to understand the impact of these programs and how various design features influence the observed results.” For example, the VBP proposal notes that “five of the seven studies did not use control groups” which is “a critical issue in evaluating the impact of a P4P program.” The VBP proposal notes that two studies with control groups “saw very modest improvements in performance” compared with improvements through public reporting, with “even less evidence of the effect of P4P on patient outcomes.” The VBP proposal also admits that the P4P programs studied “generally focused on a small set of process measures covering a handful of diagnoses” and therefore could not conclude whether “their findings generalize to other kinds of measures, such as patient experience.” With such caveats, the VBP proposal agrees with the AAMC and AHA that the evidence base for P4P is weak and requires additional research.

Hospital groups express similar concern about the use of patient-centeredness and outcomes measures that they feel require more research, experience, and field-testing to understand the ramifications fully and decrease the risk of unintended consequences. While most hospital groups recognize the validity of patient-centeredness measures, the prevailing opinion seems to be that CMS must perform more research and field testing in order to understand better how improved patient satisfaction correlates with better quality of care. For example, as Ms. Foster states, patient-centeredness scores tend to improve with new construction. Should VBP then be a vehicle to reward new construction or other amenities such as valet parking? On the other hand, if patient satisfaction is a desirable endpoint to be measured and rewarded, should it matter how hospitals achieve increases? Nevertheless, HCAHPS is a
very new program and CMS has only recently released data that investigators must analyze carefully for trends and effectiveness. Hospitals argue that patient outcomes measures are even more difficult to implement correctly given that outcomes may be beyond the hospital’s control. Therefore, outcomes measures have the greatest risk of unintended consequences, especially adverse patient selection, given that hospitals may be unwilling to admit complex and very ill patients that may harm their outcomes scores.

I suspected that providers would disagree on the VBP’s proposed methods of developing and endorsing measures, given media coverage suggesting that measure selection would be a hotly debated topic. However, the data strongly suggest that the vast majority of providers support CMS’s currently favored method of including measures adopted by the HQA and endorsed by the NQF. Given that most stakeholders approve of this method, CMS should continue this approach.

The only hospital group with a strong endorsement of VBP has been Premier. However, this may be strongly related to Premier’s role in the Premier HQID program and a financially motivated desire to continue supplying health IT and consulting services for VBP. Critics have been concerned about whether Premier’s involvement in the design and disbursement of health IT for HQID was a conflict of interest, given that Premier facilitates the acquisition of $17 billion/year worth of medical supplies for over 25% of the nation’s hospitals, and has been previously under investigation for anti-trust violations. However, these concerns did not prevent CMS from agreeing to involve Premier in data collection and analysis of HQID, and it is unknown how much involvement Premier may have in future P4P initiatives. Certainly, further involvement in VBP would be a financial windfall for Premier. Given the apparent success of Premier HQID and the credit that Premier has claimed for their role in that program, CMS will
certainly consider Premier should VBP require standardized health IT and data collection. Nevertheless, even Premier expresses concern over the VBP proposal’s assertion that no additional funding is necessary. Premier’s president, Richard Norling, has stated previously that Premier “can’t and won’t be associated with a policy that ... is funded with reductions in payments.”

The Providers: Professional Groups

Like the hospital groups, the professional groups share similar goals with those of the payer groups with respect to increasing the quality and efficiency of health care while decreasing the ballooning costs associated with Medicare. Also like the hospital groups, professional groups are concerned with the readiness of the program for full-scale adoption in Medicare, in addition to the structure of incentives and payments. Professional groups have questions about how VBP will affect their autonomy. Nowhere is this more pronounced than in the AMA’s resistance to the Medicare P4P initiatives.

Surprisingly, only the ACS, the ANA, the AONE, and the SHM represented the interests of professional groups at the Senate Finance Committee Roundtable, hardly representative of the diversity of physicians in this nation. Among the professional groups not present at the VBP Roundtable were the AMA, AAFP, ACP, and many others. While I am not aware of the selection process for stakeholders sitting at the VBP Roundtable, the Senate Finance Committee likely wished to focus more on hospital groups and inpatient providers such as hospitalists, as VBP would affect them much more directly than other physician groups, whose constituents might be highly outpatient-based and more interested in outpatient initiatives such as PGP and
PQRI. However, the fact that only two physician groups sat at the VBP Roundtable is very interesting, considering that two nursing groups were also present.

Of the physician groups at the VBP Roundtable, only the SHM strongly supports VBP. SHM’s support may be rooted in the fact that their constituents, hospitalists, stand to gain influence with the Senate Finance Committee through their support of VBP. Among the professional groups, ACS is the most vocal opponent to VBP. The ACS believes that VBP is an untested, unproven program with poorly developed and ill-defined measures and an incentive structure that has not been well-studied. The ACS shares the concerns by hospital organizations like the AHA and AAMC that VBP is not ready for implementation. These professional groups are concerned about the untested nature of P4P and would like a slow adoption of VBP together with careful field testing and substantial research concerning the ability of P4P to improve quality and efficiency of health care while also avoiding unintended consequences that may harm physician interests. Multiple professional groups also share the concern of hospital groups that with budget neutrality, VBP will primarily be a vehicle to cut costs up front rather than drive cost savings down the road through improvement in health care quality and efficiency.

Many professionals also fear the loss of autonomy through participation in VBP and another other P4P programs. Although none of the professional groups mentioned the concept of professional autonomy at the VBP Roundtable, at least one media account reports that doctors are concerned that P4P programs will erode professional autonomy and lead to "cookbook medicine." A representative of the Heritage Foundation echoes these thoughts, calling the new VBP initiatives "pay for compliance, not pay for performance," and laments that "integrity and independence of the medical profession could be compromised." The threat to physician
autonomy may be what prompted the creation of the AMA’s strict set of principles concerning any future P4P initiatives.\textsuperscript{47}

The omission of the AMA from the VBP Roundtable is particularly notable given their previous vocal opposition to certain aspects of P4P programs like VBP. Likely reasons for the AMA’s omission become clearer when we examine the relationship between the AMA and the Senate Finance Committee. In February 2008, the Senate Finance Committee banned the AMA from talks involving a bill for Medicare Reform. The AMA “lost the trust” of Senator Baucus when it “broke a confidentiality agreement about Medicare talks last year and informed state affiliates.”\textsuperscript{48} In addition, the Senate Finance Committee apparently did not appreciate that the AMA supported a budget provision in 2006 halting scheduled cuts in favor of larger future cuts, then later criticizing this provision. Overall, the AMA has been one of the most vocal critics of P4P, for better or for worse.

Despite the fact that the membership of the AMA, once the pre-eminent physician organization in the US, has now dwindled to around 24% of all US physicians, the AMA remains the largest physician group and therefore generates considerable attention with every move. In 2005, the AMA issued a set of strict principles and guidelines for any future P4P initiatives.\textsuperscript{47} At a contentious meeting with a recurring theme of needing to “stand for our principles,”\textsuperscript{49} the AMA indicated that it would oppose any plan that did not include the following principles\textsuperscript{50}:

- ensuring quality of care.
- fostering the physician/patient relationship.
- “completely” voluntary physician participation.
- accurate data and fair reporting.
- fair and equitable program incentives.

The AMA has claimed that without these principles, P4P programs could be a “lose/lose proposition for patients and their physicians” and merely benefit health insurers.\textsuperscript{51} One
physician from another specialty group remarked that the AMA’s principles seem grounded in the idea that “pay-for-performance is now seen as something that physicians need to be protected against.”

Not surprisingly, other physician interest groups have taken exception to these guidelines. They are concerned by the all-or-nothing approach that the AMA seems to be taking with P4P, fearing that their hard-line stance could compromise physicians’ bargaining position on Capitol Hill. The AMA, however, insists that its principles do allow for negotiations by starting “on principles.” Nevertheless, several specialty groups including the American Academy of Family Physicians (AAFP), the American Academy of Pediatrics (AAP), and the American College of Physicians (ACP) have indicated that they are willing to provide “support and assistance” in developing P4P programs. These specialty groups may be willing to support legislation in order to gain influence on the final design of any mandatory P4P programs, including VBP, and possibly extract other benefits in the process, such as added reimbursements.

AMA’s P4P “principles” underscore the fact that physicians are concerned about P4P and its potential consequences to physician autonomy. The primary question now remains: how much influence will physicians have over a comprehensive inpatient P4P program that will affect the vast majority of hospitals? There is some validity to the argument that the AMA’s rigid guidelines will jeopardize physician involvement in determining performance measures and other aspects of VBP. Clearly, the AMA wishes to establish ground rules while P4P is still relatively nascent in order to protect physician interests. The AMA reticence to embrace may reflect the fear that the long-term consequences of P4P implementation are not well-studied. As mentioned above, few studies have been conducted about the impact on inpatient health care quality by P4P, and the results have been mixed. Therefore, the AMA’s insistence on making
participation in any P4P program voluntary seems based on a desire to protect physicians from P4P programs that do not adequately represent physician interests.

Complicating these issues is that the AMA is no longer the monolithic voice for physician interests. In fact, as evidenced by the contentious debates leading up to the 2005 and 2007 P4P recommendations, the AMA is certainly does not present a united front. Other specialty groups, as well as many AMA constituents, fear that the AMA’s principles are unrealistic and pushing them on Capitol Hill will leave physicians with no influence over the final design of mandatory P4P programs such as VBP. Some see acquiescence to the inevitability of P4P and VBP as a method of ensuring that physicians still have some room at the bargaining table when Congress debates the details. However, while physicians generally lack the political clout that some better organized interest groups possess, one physician noted that physician resistance eventually defeated capitation, another managed care initiative. In the end, the AMA may remain influential just by staying with its principles, as well as gaining political capital with its own members and other specialty groups. The opposite is also possible: should the AMA remain so inflexible that politicians turn to other groups for feedback for P4P programs like VBP, the AMA may be headed for irrelevance.

The debate over AMA’s principles, therefore, highlights a potentially growing rift between the AMA and physician specialty groups who feel that the AMA does not represent their best interests. The AMA has been steadily losing members to these groups. The issue of P4P appears to have driven a wedge between the AMA and the other physician groups, both in the issues of AMA’s principles and in the issue of AMA’s protectiveness over its own performance measures. In fact, AMA’s actions may have increasingly united physician specialty groups against them. The AMA no longer speaks for all physicians, and a divided voice may
serve to weaken physician bargaining power in Congress and increase the likelihood that the final VBP legislation will not adequately address physician interests.

In light of the fact that most believe mandatory VBP participation for all hospitals accepting Medicare patients to be inevitable, the AMA might be better served focusing on other aspects of P4P that would prove beneficial or harmful for physicians – that is, the other four principles. For example, the concept of a mandatory P4P program that is budget neutral seems to be especially repugnant among most physicians, as it would punish a large number of hospitals and prevent cooperation that might otherwise lead to desired improvement in health care quality. Physicians are also very worried about the validity of the performance measures to be used in VBP, so the AMA would be best served trying to work toward a consensus between all physicians (including those in other specialty groups) in terms of the best performance measures to use.

Most importantly, the AMA should understand that it no longer represents the majority opinion of U.S. physicians. Its stance on P4P principles and performance measures has harmed its relationship with the other specialty groups. The AMA should strive to work with, and compromise with, other specialty groups so that its message is not blunted by physician infighting. If physicians are to have any significant political clout regarding P4P and other health care issues in Congress, the AMA must take the initiative to work with the other specialty groups in order to present a unified front.

**Other Stakeholders**

The Commonwealth Fund, the lone patient advocacy group invited to the VBP Roundtable, supports VBP. The Commonwealth Fund has stated that “quality of care does not
correspond to spending” and that “paying for high performance rewards achievement.” This, however, is not necessarily representative of patients’ views of VBP and P4P. While this paper does not focus on patients’ opinions on P4P, patients likely do not have a strong understanding or even awareness of P4P. One survey conducted in 2005 found that patients are only “modestly supportive” of any plan paying physicians more for providing better quality of care.54 According to this poll, only 38% favor the plan for paying physicians more based on achievement. However, 67% favor the plan if it lowers their costs (premiums, deductibles, and co-pays). Patients are unlikely to become better-informed about the link between improved quality and decreased costs until the issue of VBP enters the spotlight, which is unlikely as long as VBP and Medicare reform are not at the top of the national agenda.

The NQF appears to have positioned itself strongly in VBP and any future CMS P4P initiatives and programs. Given that the NQF is a non-profit organization arising from the public agenda to improve the quality of health care in the United States through setting goals and priorities for performance improvement, including standards for measurement and reporting of quality performance measures, stakeholders trust the NQF to endorse measures independent of government pressure. In addition, the NQF is staffed by a wide variety of experts and accepts feedback from a broad spectrum of stakeholders across the nation. The NQF is responsible for collecting proposed quality measures developed by stakeholders such as the AMA, endorsing and revising selected measures, and sending these measures to the CMS. The NQF will likely endorse measures for all CMS P4P programs well into the foreseeable future.
THE FUTURE OF VBP

VBP embodies the hope among all stakeholders in health care that P4P can improve quality and efficiency of health care while decreasing costs. The goals for VBP, however, differ between the payers and providers, both of which are looking to protect their own interests. VBP must represent a compromise between the payers’ desire to lower costs and providers’ desire to protect their own assets and autonomy. The government should implement VBP slowly to prevent unnecessary harm to hospitals while also ensuring that VBP is indeed buying Medicare an improvement in quality and efficiency of health care that everyone desires. In doing so, the government must also ensure that small, rural, and safety net hospitals are unharmed by providing them with special arrangements and support, if necessary.

Clearly, investigators must continue to research the ability of P4P programs to improve the quality and efficiency of care while decreasing costs, especially in the context of VBP. As shown in the mini-systematic review, the quality of evidence supporting VBP’s effectiveness in improving health care quality is mixed. Why more studies have not been performed regarding P4P, considering that VBP represents a monumental shift in the way Medicare will reimburse for inpatient care, is unclear. One possibility is that having hospitals agree to participate in P4P studies is difficult. Participation in a P4P study would require at the very least that an entire segment of the hospital to participate, which would be costly and risky as well. A hospital could face significant costs to implement required health IT. In addition, any change in health care delivery would be confusing and difficult for hospital staff. Now that the Premier HQID program has been gathering data for several years, and investigators should be able to measure changes in quality and efficiency more easily.
Interestingly enough, payers seem content to advocate for VBP despite lacking strong evidence that P4P indeed improves quality and efficiency of health care. CMS, working against a deadline of late 2007, may not have had enough time to gather sufficient data or wait for investigators to publish enough adequate studies for review. Nevertheless, the fact that the payers continue to push for VBP is strong evidence of how politically charged the issue of VBP has become. Certainly, opponents among the provider groups and lawmakers will continue to cite evidence showing that P4P does not have a strong evidence base supporting its effectiveness, while the proponents will continue to cite Premier’s data from the Premier HQID program as evidence that P4P can indeed improve the quality of health care.

Budget neutrality will continue to be a hotly debated topic. Payers such as CMS continue to believe that VBP can be effective in a budget-neutral design, while hospital groups are understandably fearful of its effects on lower-performing hospitals and collaboration between hospitals. However, political realities may force the government to implement a budget-neutral design in VBP, given the reticence of policymakers to increase taxes. One possible compromise will be to keep rewards small (1% or less of DRG payments), given that Premier HQID’s small incentives were already enough to induce change in hospital behavior. A 2-5% incentive as proposed by CMS might be too much even for hospitals to continue collaborative efforts and should be avoided at least until field-tested.

While most stakeholders may agree on the process of endorsing and selecting measures for VBP, CMS must continue to monitor the effects of its measures on hospital behavior to eliminate unintended consequences such as adverse patient selection and data gaming. Even measures as simple as requiring antibiotics within four hours for patients with community-acquired pneumonia may lead to overuse of antibiotics for unclear cases. CMS must monitor
performance closely and listen to feedback from providers, and also be willing to alter or retire measures that are not appropriate or helping to deliver an improvement in health care quality as expected. As mentioned previously, outcomes measures are at greatest risk for adverse patient selection. One method of curtailing this risk is to monitor changes in patient admissions before and after VBP implementation. CMS must also develop robust risk adjustment models to protect hospitals that normally admit complex and difficult patients from penalty.

To date, the proposal has not been brought to a vote in Congress. Many believed that the Congress needed to have passed VBP into law by late 2007 in order for the program to begin in fiscal year 2009, the deadline set by the DRA.55 However, in late 2007, Congress postponed the debate over Medicare/SCHIP reform, likely in light of looming presidential elections. As such, the immediate future of VBP is uncertain. With the uncertain climate surrounding the presidential elections of 2008, VBP will likely not be a high priority for policymakers until 2009 at the earliest.

Likely, policymakers will elect to roll out VBP slowly. Leaders in the hospital industry will continue to urge caution with VBP, believing that VBP should encourage an improvement in quality and long-term cost reductions, rather than a cut in Medicare spending. In addition, much debate and analysis remain. In the meantime, the future of VBP and P4P in Medicare, while seemingly inevitable, continues to wait for policymakers to bring the issue of quality of health care and Medicare spending back to the forefront.
Available at: http://www.senate.gov/~finance/press/Bpress/2008press/VBProundtable%20statements%20030408/Resenthal%20from


Wolter N. Statement of Nicholas Wolter and MHA - An Association of Montana Health Care Providers before the Committee on Finance of the U.S. Senate. March 6, 2008; Response by the Montana Health Care Providers to the


41. Sloane T. Everyone's seeking value; CMS payment reform plan has merit, but the debate is just beginning. Modern Healthcare, December 3, 2007; Opinions/Editorials: 18.

42. Valuck T. CMS' Progress Toward Implementing Value-Based Purchasing: April 20, 2008.


52. Romano M. Performance anxiety; is it too late for physicians to set their own rules for pay-for-performance? Modern Physician; October 1, 2005:10.


55. Becker C. High-risk proposition; With CMS' 'value-based purchasing' model, struggling hospitals are wondering if they can afford the IT needed to make it work. Modern Healthcare; December 3, 2007:6.
# Appendix 1: Glossary

<table>
<thead>
<tr>
<th>Acronym/Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAFP</td>
<td>American Academy of Family Physicians</td>
</tr>
<tr>
<td>AAMC</td>
<td>The Association of American Medical Colleges</td>
</tr>
<tr>
<td>AAP</td>
<td>American Academy of Pediatrics</td>
</tr>
<tr>
<td>ACP</td>
<td>American College of Physicians</td>
</tr>
<tr>
<td>ACS</td>
<td>American College of Surgeons</td>
</tr>
<tr>
<td>AHA</td>
<td>American Hospital Association</td>
</tr>
<tr>
<td>AMA</td>
<td>American Medical Association</td>
</tr>
<tr>
<td>AMI</td>
<td>acute myocardial infarction</td>
</tr>
<tr>
<td>ANA</td>
<td>American Nursing Association</td>
</tr>
<tr>
<td>AONE</td>
<td>American Organization of Nurse Executives</td>
</tr>
<tr>
<td>BCBSA</td>
<td>Blue Cross Blue Shield Association</td>
</tr>
<tr>
<td>Benchmark</td>
<td>A reference point or basis of comparison.</td>
</tr>
<tr>
<td>CABG</td>
<td>Coronary Artery Bypass Graft</td>
</tr>
<tr>
<td>CAP</td>
<td>Community-acquired pneumonia</td>
</tr>
<tr>
<td>Composite</td>
<td>An aggregation of individual measures.</td>
</tr>
<tr>
<td>CMS</td>
<td>Centers for Medicare and Medicaid Services</td>
</tr>
<tr>
<td>DHHS</td>
<td>Department of Health and Human Services</td>
</tr>
<tr>
<td>DRA</td>
<td>Deficit Reduction Act of 2005</td>
</tr>
<tr>
<td>DRG</td>
<td>Diagnosis-Related Group, related to Medicare reimbursement.</td>
</tr>
<tr>
<td>EMR</td>
<td>Electronic Medical Records</td>
</tr>
<tr>
<td>FAH</td>
<td>Fee-For-Service, or a method of reimbursement that pays providers based on each service rendered.</td>
</tr>
<tr>
<td>FFS</td>
<td>Fee-For-Service, or a method of reimbursement that pays providers based on each service rendered.</td>
</tr>
<tr>
<td>HbA1c</td>
<td>Hemoglobin A1c, or glycosylated hemoglobin, is a measure of chronic diabetes.</td>
</tr>
<tr>
<td>HCAHPS</td>
<td>A standardized survey instrument and data collection methodology for measuring patients’ perspectives on hospital care.</td>
</tr>
<tr>
<td>Hospital Comp</td>
<td>A tool on the CMS website providing information on hospital performance of care on certain measures: <a href="http://www.hospitalcompare.hhs.gov">www.hospitalcompare.hhs.gov</a></td>
</tr>
<tr>
<td>HF</td>
<td>Heart Failure</td>
</tr>
<tr>
<td>HQA</td>
<td>Hospital Quality Alliance</td>
</tr>
<tr>
<td>HQID</td>
<td>Hospital Quality Incentive Demonstration, associated with Premier.</td>
</tr>
<tr>
<td>IOM</td>
<td>Institute of Medicine</td>
</tr>
<tr>
<td>IT</td>
<td>Information technology</td>
</tr>
<tr>
<td>IPPS</td>
<td>Inpatient Prospective Payment System</td>
</tr>
<tr>
<td>LDL-C</td>
<td>a type of cholesterol</td>
</tr>
<tr>
<td>MMA</td>
<td>Medicare Modernization Act of 2003</td>
</tr>
<tr>
<td>NQF</td>
<td>National Quality Forum, a semi-public organization focused on improving quality that represents many interests in health care</td>
</tr>
<tr>
<td>P4P</td>
<td>Pay-for-performance; originally a business payment model that reimburses</td>
</tr>
<tr>
<td>Acronym/Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>based on results attained.</td>
<td></td>
</tr>
<tr>
<td>P4R</td>
<td>Pay-for-reporting, or reimbursement for reporting data.</td>
</tr>
<tr>
<td>RHQDAPU</td>
<td>Reporting Hospital Quality Data for Annual Payment Update, or CMS's first nationwide pay-for-reporting program.</td>
</tr>
<tr>
<td>Risk adjustment</td>
<td>A method of reducing the effects of characteristics of patient population outside the control of providers on performance measures.</td>
</tr>
<tr>
<td>SCIP</td>
<td>Surgical Care Improvement Project</td>
</tr>
<tr>
<td>SHM</td>
<td>Society of Hospital Medicine</td>
</tr>
<tr>
<td>VBP</td>
<td>Value-Based Purchasing, or a CMS policy mechanism that links payment to performance with regards to Medicare</td>
</tr>
</tbody>
</table>
### Appendix 2: P4P literature

<table>
<thead>
<tr>
<th>Study</th>
<th>intervention group</th>
<th>contemporary control</th>
<th>measures</th>
<th>time period</th>
<th>summary of findings</th>
<th>conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutler et al.</td>
<td>165 pts in Mercy Medical Group, with diabetes and elevated LDL-C, treated under P4P program</td>
<td>1694 pts in MMG with similar conditions under routine care</td>
<td>LDL-C testing, LDL-C control</td>
<td>2 years</td>
<td>75th and 90th percentiles for lab testing and goal attainment</td>
<td>if these rates were extended to entire P4P population, medical group would have received substantial incentive payments</td>
</tr>
<tr>
<td>Doran et al.</td>
<td>8105 family practices in England</td>
<td>none</td>
<td>indicators for asthma, CA, CHD, COPD, DM</td>
<td>1 year</td>
<td>achievement 81-96% on all clinical indicators</td>
<td>high levels of achievement suggest that targets were too easy to achieve</td>
</tr>
<tr>
<td>Glickman et al.</td>
<td>54 hospitals in Premier HQID</td>
<td>446 hospitals</td>
<td>use of ACC/AHA class I guideline recommended therapies, in-hospital mortality</td>
<td>3 years</td>
<td>no significant difference in quality of care or outcomes for acute MI</td>
<td>Participation in Premier neither improved or worsened quality of care or outcomes for acute MI.</td>
</tr>
<tr>
<td>Grossbart et al.</td>
<td>4 acute care hospitals in Catholic Healthcare Partners, participating in Premier HQID</td>
<td>6 hospitals in Catholic Healthcare Partners, declining to participate in Premier HQID</td>
<td>AMI, HF, pneumonia</td>
<td>2 years</td>
<td>Significant difference in composite quality score of P4P participants (9.3%) compared with control (6.7%), due mainly to HF measures (19.2% versus 10.9%)</td>
<td>Premier HQID effective in improving quality of healthcare especially in HF care.</td>
</tr>
<tr>
<td>Lindenauer et al.</td>
<td>207 hospitals participating in Premier HQID</td>
<td>406 hospitals with public reporting only</td>
<td>10 individual, 4 composite measures</td>
<td>2 years</td>
<td>P4P participants improve 2.6-4.1% over 2 years</td>
<td>Hospitals in intervention group attained modest improvements.</td>
</tr>
<tr>
<td>Mandel et al.</td>
<td>43 practices</td>
<td>none</td>
<td>flu shot %, controller med % for persistent asthma, written self-management plan %</td>
<td>3 years</td>
<td>% receiving perfect care from 4% to 88%; % receiving flu vaccine from 22% to 62%</td>
<td>P4P can help drive improvement in quality of care, but system changes required as well.</td>
</tr>
<tr>
<td>Premier White Paper: Year 1</td>
<td>262 hospitals participating in Premier HQID</td>
<td>none</td>
<td>measures associated with AMI, CABG, CAP, HF, hip/knee replacement</td>
<td>1 year</td>
<td>improvements ranging from 3.4% in AMI to 9.8% in CAP from beginning to end of year 1</td>
<td>no conclusion provided</td>
</tr>
<tr>
<td>Premier White Paper: Year 2</td>
<td>253 hospitals participating in Premier HQID</td>
<td>none</td>
<td>measures associated with AMI, CABG, CAP, HF, hip/knee replacement</td>
<td>1 year</td>
<td>improvements ranging from 3.5% in AMI to 10.4% in HF from year 1 (mean) to year 2 (mean)</td>
<td>no conclusion provided</td>
</tr>
<tr>
<td>Rosenthal et al.</td>
<td>California physicians group</td>
<td>Pacific Northwest physicians group</td>
<td>cervical cancer screening, mammography, HbA1c testing</td>
<td>1 year</td>
<td>improvements for cervical cancer: 5.3% v. 1.7% (control); mammography, 1.9% v. 2.2% (control); HbA1c, 2.1% v. 2.1%; $3.4 million spent in bonuses</td>
<td>little quality gain for money, mainly rewards providers with higher baseline performance</td>
</tr>
</tbody>
</table>
Appendix 3: Stakeholder Questions

The following is a list of questions used for all stakeholder interviews.

First of all, would you say Medicare is a logical place to try to implement P4P?

Who would you say are the main advocates and opponents for P4P?

My next question is about physician buy-in. I got the impression that some are strong advocates while others have great concern. What are your impressions of physicians' views in general?

Some people argue that P4P can control costs while improving quality. Do you think it has that potential?

Other people are concerned that P4P programs will do more harm than good. For example, some believe it may have the effect of adverse selection on patients and lead to such things as gaming. What do you think about that?

Are the benchmarks CMS has now good enough to let us measure whether P4P programs really can improve quality?

Do you think electronic medical records or other health IT are required for P4P?

Do you believe transparency is a good idea for P4P?

Should patient perception of quality play a role in P4P programs?

What about smaller and rural hospitals? Do P4P programs pose any challenges or advantages to them?
What about budget neutrality? Can P4P work with budget neutrality?

How do you think budget neutrality could affect collaboration between hospitals?

Recently, as you know, CMS has said that it will no longer reimburse for so-called "never" events [explain if necessary: foreign objects remaining in the body, amputation of the wrong limb, decubiti, death from medical error, etc.]. Do you think this will make a significant difference to safety? Or will it produce less than people might expect? How do you think it will play out?

Let me move to the specific questions about the Premier demonstration and the VBP proposal. What goals was the Premier demonstration meant to reach, and how did it do?

Do you think we have enough evidence to adopt VBP?

What do you see as the future of VBP? Do you expect it to be national policy in the near future?