



# The California Pay For Performance Program

The Second Chapter  
Measurement Years 2006 - 2009

An illustration of two men in a meeting room. One man in a white shirt stands with his back to the viewer, arms outstretched, pointing towards a large whiteboard on the left. The other man, also in a white shirt, stands on the right, holding a large red block. The floor is covered with various colored blocks (purple, orange, green, blue) arranged in a structure. The background is a textured green wall.

Innovation Through Collaboration

June 2009

## **ABOUT THE INTEGRATED HEALTHCARE ASSOCIATION (IHA)**

The Integrated Healthcare Association (IHA) is a statewide leadership group that promotes quality improvement, accountability, and affordability of health care in California. The IHA P4P program is the largest non-governmental physician incentive program in the U.S. It includes 8 health plans and over 200 medical groups representing 35,000 physicians providing care for 11.5 million HMO members. Other IHA programs include: medical technology value assessment and purchasing; the measurement and reward of healthcare efficiency; and healthcare affordability.

### **Credits**

Editors: Tom Williams and Dolores Yanagihara

Contributing Editors: Steve McDermott and Leslie Rose

## THE CALIFORNIA PAY FOR PERFORMANCE PROGRAM THE SECOND CHAPTER

Measurement Years 2006 – 2009



June 2009

The California Pay for Performance program completed its sixth year in 2008. All of the original stakeholders continue to participate actively in the program, despite many challenges and differences of opinion along the way. This includes the following California health plans: Aetna, CIGNA, Anthem Blue Cross, Blue Shield, Health Net, Kaiser Permanente, UnitedHealthcare (formerly PacifiCare), and Western Health Advantage. It also includes over 200 participating physician groups, representing 35,000 physicians, who continue to demonstrate their commitment to improve the quality and affordability of healthcare.

Many organizations and individuals have contributed their time and expertise to developing and maintaining the California Pay for Performance program. The California HealthCare Foundation (CHCF) has provided grant funding, leadership, and encouragement throughout the process. The National Committee for Quality Assurance (NCQA) has been a business partner since the program's inception, providing reliable data collection, reporting, and hands-on measurement expertise. The Pacific Business Group on Health (PBGH) has provided technical support and leadership both within the program committee structure and externally. The California Association of Physician Groups (CAPG) has helped navigate both political and administrative challenges.

The program governance involves numerous committees, including some highly technical in nature. The individuals currently serving on these important committees are listed in Appendix A. They and their predecessors have committed thousands of hours to reviewing documents, attending meetings and at times patiently enduring our consensus decision-making process.

We offer our heartfelt appreciation and thanks to all these organizations and individuals. You have helped build a foundation for performance measurement that extends beyond the borders of California.

A handwritten signature in black ink that reads "Donald G. Rebhun".

Don Rebhun, M.D.  
Board Chair, Integrated Healthcare Association  
Regional Medical Director, HealthCare Partners

A handwritten signature in black ink that reads "Tom Williams".

Tom Williams, M.P.H., M.B.A.  
Executive Director  
Integrated Healthcare Association

## TABLE OF CONTENTS

<b>I. Introduction and Overview</b>	<b>3</b>
<b>II. Measurement Set</b>	<b>6</b>
<b>III. Performance Results</b>	<b>9</b>
<b>IV. The “Pay” in Pay for Performance</b>	<b>15</b>
<b>V. Public Reporting and Public Recognition</b>	<b>18</b>
<b>VI. Program Management and Organization</b>	<b>20</b>
<b>VII. Lessons Learned</b>	<b>22</b>
<b>VIII. Looking Ahead</b>	<b>25</b>
<b>IX. Appendices</b>	<b>26</b>
<b>References</b>	<b>32</b>

## I. INTRODUCTION AND OVERVIEW

### Introduction

The first Integrated Healthcare Association (IHA) white paper, *“Advancing Quality through Collaboration: the California Pay for Performance Program”* was published in February 2006. The report was issued in conjunction with the first National Pay for Performance Summit held in Los Angeles, California, where one speaker was moved to announce the event as the official marking of the pay for performance (P4P) “honeymoon” period. This subsequent IHA white paper entitled *“The Second Chapter”* describes the more recent phase of the California P4P program, including its post-honeymoon successes, disappointments and growing pains.

The California P4P program finds itself at a crossroads. A foundation of trust built upon reliable performance measurement and reporting has been established and physician groups have embraced quality as a strategic imperative. Yet the enthusiasm of participating health plans has waned as Health Effectiveness Data and Information Set (HEDIS) scores have not advanced their national quality rankings and pressures to moderate premium increases have overwhelmed other concerns.

Nationally, P4P continues to expand, even as research tells us that the results are mixed.<sup>1,2</sup> As we debate our progress, let us not forget how recently the platitude “You cannot measure quality in healthcare” was common wisdom. If nothing else, P4P has helped dispel that once strongly held belief. However, the promise of P4P has not been fully realized. With that in mind, we share our results, and describe our efforts to advance P4P.

### KEY POINTS:

- Creation of a single performance-based measure set and public report card was a major step forward.
- The use of aggregated data from multiple payers to score results significantly increased measurement reliability and the trust of physicians.
- P4P has created a collaborative environment between health plans and physician groups, which serves as a platform for other initiatives.
- Physician groups have accepted the challenge of performance measurement and actively engaged in quality improvement efforts.
- Steady, incremental quality improvements have been realized, but breakthrough improvement has not been achieved.
- Dramatic regional/geographic variations in quality have surfaced.
- Payments represent a small percentage of compensation relative to other programs, which likely contributed to program performance.
- Wide variability in payments by health plans raised concerns about “free riders” motivating plans with higher payments to reduce payment levels.

### Overview

The California P4P program is a story of collaboration, consensus building and engagement. It began in 2003 with six health plans, 200 physician groups and almost 7 million commercial HMO enrollees, using a uniform measure set of 25 measures in three domains (clinical, patient experience and information technology). Today, this collaboration has grown to include eight health plans with 11.5 million commercial HMO enrollees and a set of 68 measures in five domains (clinical, coordinated diabetes care, patient experience, IT-enabled systemness, and efficiency). Performance results are publicly reported and IHA publicly recognizes the state’s top performers, as well as those most improved.

Time, feedback, evaluation and introspection have allowed us to draw conclusions about P4P. The value of collaboration among health plans and physician groups to create and use uniform measures cannot be overstated. Physician groups consistently report that the P4P program has allowed them to focus on quality improvement in an environment of limited resources.

The aggregation of multi-payer and physician group data to determine a single “score” is far superior to individual health plan scoring, which uses only a slice of a physician group’s patient population.<sup>3</sup> The results are significantly more reliable, and perhaps just as important, more trusted. Scores derived from aggregated data are used for both payment and public reporting.

*“The major achievement is bringing large health plans, medical groups, NCQA and PBGH together to produce a statewide state-of-art quality report card based on best of class features.”*

–Michael Belman, M.D., M.P.H., F.A.C.P., Medical Director, Quality and Innovation, Anthem Blue Cross of California; IHA P4P Steering Committee Member

Collaboration and consensus building have meant involving those being measured in the design of the measure set. This enhances P4P face validity, essential to physician engagement. These efforts have generated the intended response from physician groups: acceptance of measurement and engagement in improvement. Evaluation underscores the result: interviews by program evaluators in 2005 and again in 2007 demonstrate that physician groups participating in P4P have increased their focus on quality improvement activities. There is now widespread support for payment incentives, clinical IT adoption, and data collection for quality management. Overall, physician groups gave P4P an average rating of 4 on a scale of 1 to 5, with higher scores indicating a more favorable response.<sup>4</sup>

*“P4P raised awareness and subsequent acceptance for objective measures for quality performance assessment. It has also contributed to cultural change – increased accountability, behavioral changes and transparency.”*

–Jerry Penso, M.D., Associate Medical Director, Quality Programs, Sharp Rees-Stealy Medical Group; IHA P4P Executive Committee Member

Clinical quality performance on individual measures has improved from 5.1 to 12.4 percentage points since the inception of the program. This is understandably consistent with national trends, because most HMO plans across the country participate in P4P programs.<sup>5</sup> Ratings on patient experience increased only marginally – a major disappointment. The greatest improvement involved physician group use of information technology. There was a 7 percent average annual increase in groups adopting specific IT capabilities, including population management and point-of-care activities. Not surprisingly, IT adoption correlated with better clinical performance.

Performance results also revealed a large geographic variation, seemingly along socioeconomic lines. This dynamic is now being studied. We expect special initiatives will be needed among the groups in lower performing regions to raise overall state performance scores.

While overall results are encouraging, they do not demonstrate widespread breakthroughs or improvements; therefore, health plans are questioning their return on investment for incentive payments. Interviews by program evaluators reflect waning health plan engagement in the currently structured P4P program, and the need for fundamental program change. Overall, health plans gave P4P an average rating of just 2.5 on a scale of 1 to 5.<sup>6</sup>

*“The greatest disappointment is that the program has not demonstrated significant breakthroughs in health care quality and efficiency. Despite over \$260 million of incentive payments collectively paid by health plans on the IHA P4P measures, it has failed to demonstrate substantial overall improvement.”*

–Sam Ho, M.D., EVP and Chief Medical Officer,  
UnitedHealthcare; IHA P4P Steering Committee Member

Despite the expansion and growing sophistication of the measure set, payment has not kept pace. Overall payment amounts declined for the first time in 2008, in part due to a shift in budget by one plan to gain-sharing outside of P4P. Health plans have communicated the need to incorporate appropriate use and efficiency measures to rationalize higher incentive payment amounts. In response, an initial set of appropriate use measures has been added to the measure set, along with a recommended gain-sharing approach to share cost savings with physician groups.

Incentive payments from different health plans have been variable in both relative amount and methodology. Program goals call for standardizing methodology. Recommendations to create greater uniformity in payment methods include an increased emphasis on paying for improvement.

*“Looking back it seems obvious, but there was unequal attention to measurement and payment. While the total amount of incentive dollars paid seemed high, they were at best about 2% of total revenues to physician groups, insufficient to generate breakthrough improvement. Disappointing to me was the failure of some health plans to support the program financially.”*

–Steve McDermott, CEO, Hill Physicians Medical Group;  
IHA P4P Executive Committee Chair

Looking forward, fundamental program changes are needed to keep P4P relevant and sustainable. An essential ingredient is balancing quality and cost efficiency improvement with higher rewards. In the current economic environment, increased funding appears daunting. A common vision is emerging around the notion of enhanced gain-sharing – funding increased incentives through cost savings – embedded in performance-based contracts.

*“To drive major improvements, performance-based payments must exceed 10% of total provider income. Incentives of this magnitude can only be mobilized if they originate in payer savings.”*

–Arnold Milstein, M.D., M.P.H., U.S. Healthcare Thought Leader, Mercer Health & Benefits; Medical Director, Pacific Business Group on Health; IHA P4P Steering Committee Chair

Linking performance to differential benefit design and member cost sharing is also under consideration. As health plans and physician groups concur on appropriate measures of quality and cost, it is conceivable they might collaborate on common definitions of value. Introducing differential member co-payments, coinsurance or premium sharing based upon these definitions might introduce an additional motivation for behavior changes among providers and patients.

*“We must introduce mechanisms that reward consumers who select more efficient healthcare providers. This can only be accomplished on a foundation of valid measures; reliable measurement and differential benefit design that rewards efficient healthcare providers with more patients.”*

–Alain Enthoven, Ph.D., Marriner S. Eccles Professor of Public and Private Management, Graduate School of Business, Stanford University; IHA P4P Payment Committee Chair

## II. MEASUREMENT SET

A central goal established at the start of the California Pay for Performance (P4P) program was to develop a comprehensive measurement set to drive breakthroughs in performance. The initial measurement set began with three domains and a total of 25 measures in the areas of clinical, patient experience and information technology (IT) adoption to support patient care. Since that time, the number of measures has more than doubled, as illustrated by Figure 1.

**Figure 1 – Measurement Set Development 2003 – 2009**

Measurements	2003	2008	2009
Clinical – Preventive	8	11	13
Clinical – Chronic	3	6	6
Clinical – Acute	0	4	4
Patient Experience	6	9	9
Information Technology (IT)	8	11	11
Systemness	0	8	8
Coordinated Diabetes Care	0	10	11
Efficiency / Resource Use	0	1	6
<b>Total</b>	<b>25</b>	<b>60</b>	<b>68</b>

Not only has the number of measures dramatically increased, the measurement set has become more sophisticated. For measurement year 2009, there are five measurement domains (four of which are eligible for bonus incentive payments and one for gain-sharing) and 68 measures. The most significant additions include a new coordinated diabetes care domain, measures of systematic care processes, and efficiency measures. Weighting for each domain determines how much of the incentive payment budget is allocated to that domain. Although domain weighting has been modified each year based on changes to the measurement set and program priorities, it still favors clinical performance measurement.

The measurement domains and payment weights for measurement year 2009 are as follows:

- **Clinical quality (40%):** includes preventive, chronic, and acute care. Incorporates process

and outcomes measures, using standardized national measures wherever possible.

- **Patient experience (20%):** measures patient ratings of care received from their doctors and other providers in their physician group.
- **IT-enabled systemness (20%):** evaluates support and infrastructure that physician groups use for systematic processes of care. This includes population management, point-of-care activities, care management processes, and individual physician-level measurement and incentives.
- **Coordinated diabetes care (20%):** promotes efforts to redesign processes and create a systematic approach to diabetes care. Measures in this domain are all diabetes-related and include both process and outcome clinical measures; population management activities such as registries, actionable reports and individual physician-level measurement; and care management processes.
- **Resource use and efficiency (gain-sharing – separate from incentive pool):** assesses use of key health care services to identify variation and maximize limited resources.

### Clinical Quality Measurement

Integrated Healthcare Association’s development of clinical performance measures started with underused clinical process measures. Standardized HEDIS measures that health plans were already calculating and reporting at a plan level were adapted for physician group level measurement. Over time, P4P has incorporated some overuse/misuse measures and outcomes measures, still sourced primarily from existing HEDIS measures. Going forward, the goal is to add more outcomes and specialty measures.

The use of outcomes measures has always been considered a critical component of the evolving P4P measurement set. There are currently two clinical outcomes measures: HbA1c control (poor control and good control) and LDL control for people with diabetes and cardiovascular disease. Adoption of other outcomes measures, such as blood pressure control, is limited by the electronic

data currently available. As adoption of electronic medical records (EMR), disease registries, and/or CPT-II codes increases, the scope of potential measures is expected to expand.

To expand the measurement set, IHA has had an ongoing discussion about innovation versus use of national standardized measures. In February 2008, the P4P Steering Committee adopted a Measure Selection Framework, which specified that P4P should adopt measures that are nationally vetted, standardized, and endorsed by the National Quality Forum (NQF). However, only a limited number of NQF-endorsed ambulatory care measures apply to physician group measurement utilizing only electronic data.

For example, in an attempt to incorporate NQF measures, the National Committee for Quality Assurance (NCQA) completed a review for IHA of the 499 NQF National Voluntary Consensus Standards endorsed as of October 2008.<sup>7</sup> They searched for ambulatory care setting, using only electronic claims data, and removed Medicare-only measures. Of the original 499, only 71 measures were germane. Upon a detailed review of these measures, NCQA further determined that many of the measures were already included in the current P4P measurement set. After removing all current P4P measures, there were 40 remaining NQF-endorsed measures. However, many of these measures appeared to require medical record review even though the methodology listed electronic claims. Additionally, some of the measures did not meet IHA measure selection criteria for relevance to a significant population. That left a handful of measures, which all had either small denominator or data collection issues due to the carve-out of behavioral health services in California. Based on these findings, the P4P program determined that there were no other potential NQF-endorsed measures for P4P adoption at that time.

### **Patient Experience Measurement**

Patient experience is measured via a survey administered by the Pacific Business Group on Health. Since the inclusion of patient experience in the P4P measurement set, physician group

participation in the survey has increased sharply. In 2002, approximately 80 physician groups participated in the survey. Prompted by P4P, use of the survey tool for assessing patient experience grew to 177 physician groups by 2008. During this time, the survey tool was transitioned to the nationally vetted, standardized clinician and group Consumer Assessment of Healthcare Provider and Systems (CG-CAHPS) survey. Use of an established survey tool is efficient and allows comparison of results over time and across locations.

Patient experience was initially measured in four areas: communication with doctors, overall rating of care, specialty care, and timely access to care. Care coordination was added in 2005, and two additional areas were added in 2008: helpful office staff and health promotion.

### **Information Technology and “Systemness” Measurement**

At the development stage of P4P there were no plans to include information technology (IT) in the measurement set. This was reconsidered based on the belief that investment in IT infrastructure was essential for achieving breakthroughs in clinical performance.

The initial IT domain required physician groups to show capabilities in one of two domains to earn incentive payment: clinical data integration at the group level and clinical decision support at the point-of-care. This includes population management activities, such as patient registries and actionable reports, and point-of-care activities, such as e-prescribing, electronic checks for drug interaction, accessing clinical results, and retrieval of lab results.

In measurement year 2007, the IT domain was expanded to the IT-Enabled Systemness domain. The original IT measures were maintained, and measures were added to assess whether or not a group has systematic processes for identifying and acting on patients in need of additional services due to a chronic condition or hospitalization. The expanded IT domain also considers whether or not a group measures and provides incentives for

individual physician performance by utilizing evidence-based metrics. This expansion was prompted by the observed link between IT capabilities and clinical performance, and the ongoing desire to reach breakthrough improvement.

### **Coordinated Diabetes Care Measurement**

The Coordinated Diabetes Care domain was developed using existing measures for the management of diabetes in combination with new process of care measures. The intention of this domain is to promote process redesign in order to systematize diabetes care. Once a coordinated care system is created, it can easily be expanded to include other chronic conditions.

### **Efficiency Measurement**

The healthcare market has changed considerably since the start of P4P. Purchasers are demanding more moderate premium increases and health plans have insisted that P4P incorporate measures and payment incentives that reward cost efficiency. Initially, traditional utilization measures, such as inpatient days per 1000, were considered, but P4P stakeholders decided that a more sophisticated approach using episode-of-care measures offered greater opportunities.

IHA proceeded to test episode-of-care measures using aggregated health plan claim data. Two challenges emerged. First, claim data for capitated professional services collected in the form of encounters was incomplete for certain health plans and physician groups. Second, sample size, even with the physician group as the unit of measure, was often insufficient. As it became clear these issues could not be easily or quickly resolved, IHA introduced population-based appropriate resource use measures for its efficiency measure domain.

Appropriate resource use measures have been added to the P4P measure set for measurement year 2009. The recommended payment methodology for these measures is gain-sharing. These types of measures and incentives have been used at most health plans in some form, but have not been standardized across plans. Bringing them into P4P establishes common definitions, specifications, and a process to aggregate results across plans.

The complete 2009 Measurement Set is noted in Appendix B.

### III. PERFORMANCE RESULTS

Since program inception, physician groups have improved in all measurement areas: clinical performance improved by an average of 3 percentage points annually and patient experience saw an initial jump of 2.2 percentage points and then leveled off. The greatest improvement has come in physician group use of information technology, with an average annual increase of 7% in the number of groups adopting specific IT activities. These physician group improvements are laudable. Results for each of these three initial measurement domains are detailed below.

#### Clinical Quality Results

##### Steady Clinical Quality Improvements

Statewide the average clinical performance in P4P has steadily improved each year across all physician groups. Performance on individual clinical measures has improved from 5.1 to 12.4 percentage points since the inception of the measure. The average improvement across all clinical measures is comparable with rates of improvement nationally. Figure 2 shows aggregated P4P performance for baseline compared to the last two years of performance for select clinical measures.

**Figure 2 – California P4P Clinical Results (Baseline to 2007)**

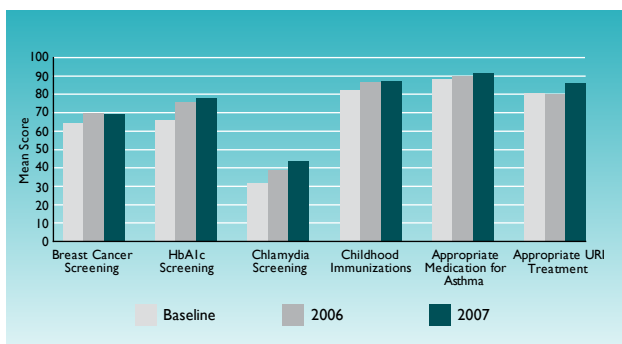
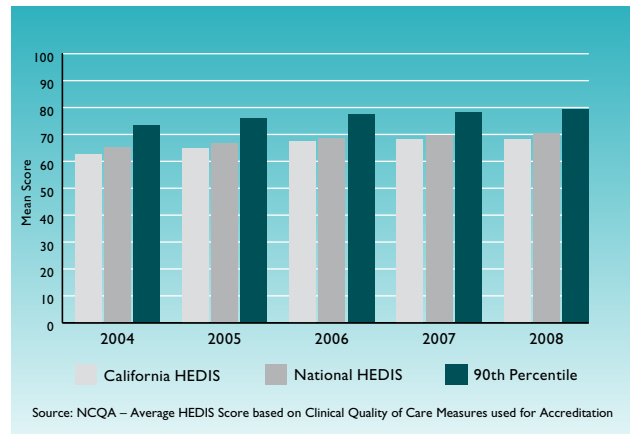


Figure 3 compares California health plan HEDIS performance to the national average and the national 90th percentile. Closing the gap between California and national performance remains a primary objective of the program.

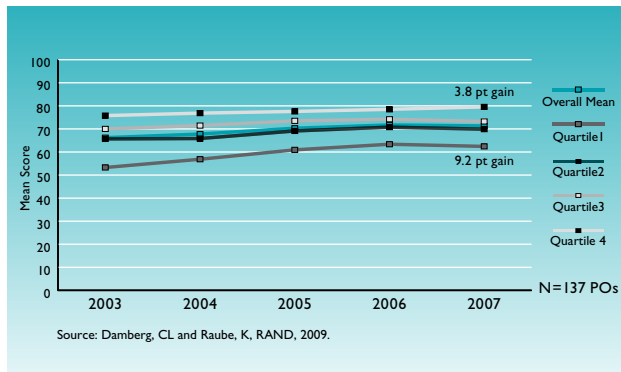
**Figure 3 – California versus Nation: Health Plan HEDIS Performance (2003-2007)**



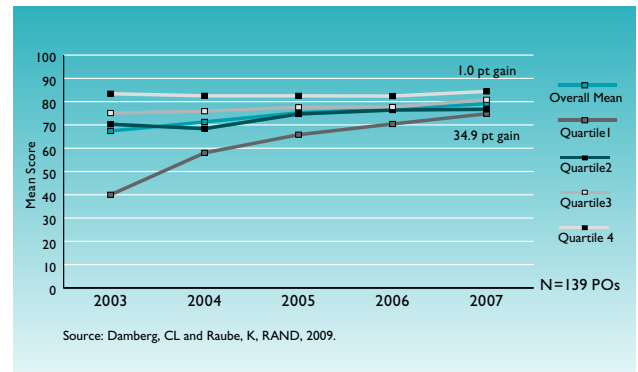
##### Striking Differences in Clinical Improvement by Performance Quartile

Overall clinical performance improvement has been incremental, but dramatic differences in clinical performance improvement emerge when looking at performance quartiles. Physician groups that started in the lowest performance quartile improved much more significantly in terms of absolute improvement than physician groups that started in the top performance quartile. Depending on the measure, the lowest quartile performers have improved from twofold to many times more than the top quartile performers. Figures 4 and 5 depict these differences. This pattern of improvement has caused substantial reductions in the variation across groups. However, the lower performers are generally smaller physician groups, and they do not account for a large percent of the population. Thus, this reduction in variation has not necessarily “moved the dial” when considering performance rates for the whole population.

**Figure 4 – Breast Cancer Screening Improvement by Performance Quartile (2003-2007)**



**Figure 5 – Diabetes HbA1c Screening Improvement by Performance Quartile (2003-2007)**



## SHARP REES – STEALY MEDICAL GROUP: A TOP PERFORMER ALSO EARNS MOST IMPROVED AWARD

### Who

Sharp Rees-Stealy Medical Group is a multispecialty medical group with 18 locations throughout San Diego, and is part of Sharp HealthCare, a not-for-profit healthcare delivery system.

### What

Sharp Rees-Stealy Medical Group has been recognized as a P4P top performer for the past several years and they have improved in every single clinical and patient experience measure. In 2008, they were recognized both as “top performer” in the state and “most improved” within their region.

### How

Sharp Rees-Stealy and all of Sharp HealthCare already had in place an initiative focused on the organization’s culture and interaction with patients.

To ensure continuous improvement, Sharp Rees-Stealy formed a P4P Action Team with employees and physicians from across departments including nursing, radiology, lab, patient contact center, and IT. In one example of care redesign, Sharp Rees-Stealy now offers mammography services as a same day appointment compared to the previous longer wait time for appointments.

Accountability is also critical to Sharp Rees-Stealy’s P4P and every employee’s performance targets are linked to the achievement of goals related to quality, service, people, finance, growth and community. Physician performance is also closely aligned with achievement of key targets.

Each year, Sharp Rees-Stealy selects key strategic quality measures. Most recently they focused on LDL control and diabetes blood sugar control. The group significantly improved in both measures even while starting at an already high level of performance. Efforts that led to continued improvement included: setting a stretch goal and providing tools to improve performance, generating quarterly performance reports on individual physicians, supplying actionable patient lists, developing internal prompts, making pre-appointment calls, developing a post-appointment process and adding diabetes care managers.

### Significant Regional Variation in Clinical Results

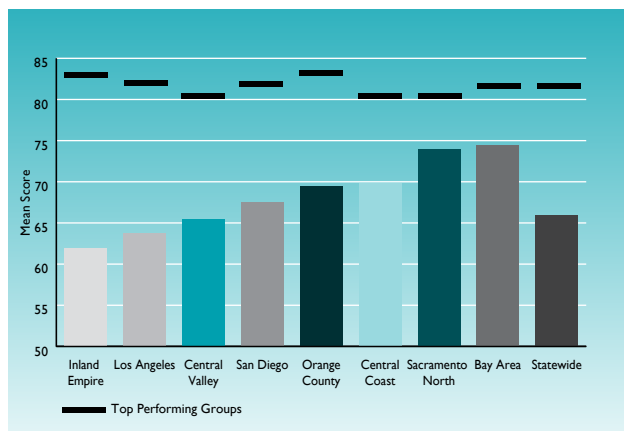
Applying the HEDIS-based clinical measures to the physician group as the unit of measure has exposed striking geographic performance variations within the state for clinical quality performance measures. Figure 6 highlights performance variation geographically. When combining clinical performance into a single clinical composite score, physician groups in the Inland Empire region achieved an average clinical composite score of 61%, versus an average clinical composite score of 74% for physician groups in the highest performing region, the Bay Area. This geographic variation helps explain the overall mediocre performance of California, as compared to other states, despite California’s many well recognized physician groups and institutions.

The geographic disparity in performance results calls for special attention to low performing areas of the state. This information has inspired collaboration among health plans, physician groups, state government and others.

Performance measurement at the physician group level also revealed the ability of some groups to perform at the highest level throughout the state.

*The results suggest that well organized physician groups with strong management, medical leadership, information technology and the ability to allocate capital across geographic areas are capable of achieving high levels of performance, even in those regions of the state with overall low performance.*

**Figure 6 – Clinical Regional Performance Variation (2007) Clinical Composite Score**



### Patient Experience Results

#### Marginal Patient Experience Improvement

Patient experience performance improvement looked promising initially, with a 2.23 percentage point increase from 2003 to 2004 in the average performance across all patient experience measures for all physician groups participating in the patient experience survey. The average scores from 2004 to 2005 also increased, but the increases were not comparable since there was a change in survey tool and methodology. From 2005 onward, patient experience performance improvement has been waning. The average improvement from 2005 to 2006 was 0.4 percentage points, and from 2006 to 2007, scores decreased by an average of 0.14 percentage points. Figure 7 shows these year-over-year changes in patient experience performance, based on the unweighted average score across all P4P patient experience measures for all physician groups that participated in the patient experience survey for that year.

**Figure 7 – California P4P Year-Over-Year Patient Experience Results**

YEAR	AVERAGE PERCENTAGE POINT IMPROVEMENT
2003-2004	2.23
2004-2005	Methodology change – not comparable
2005-2006	0.40
2006-2007	-0.14

Across all years, patients rated their doctor-patient interactions and experiences with their personal doctor, specialist and general health care significantly better than measures related to access to care, care coordination and getting an appointment with a specialist.

**Regional Variation in Patient Experience Performance**

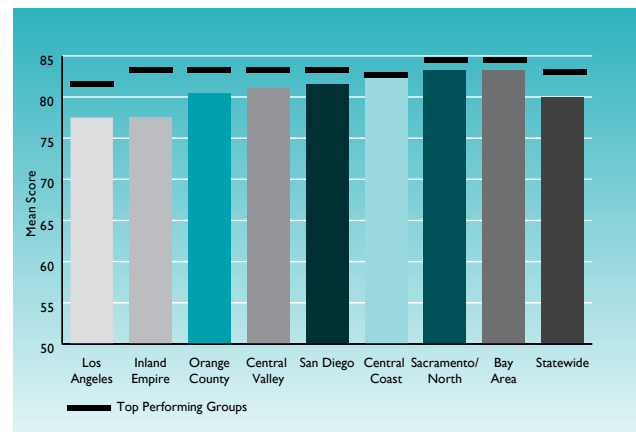
Regional variation was also observed in patient experience performance, although the performance gap was smaller: 78% composite score for the Inland Empire and Los Angeles regions, versus an 83% composite score for both the Bay Area and Sacramento/North regions.

*It is worth noting that the regional scores for patient experience follow the same general pattern as the regional scores for clinical quality.*

Patient experience is an area in which incentives alone have not demonstrated a significant effect, despite concerted efforts by many physician groups. IHA stakeholders have realized that incentives in combination with quality improvement, or “learning” collaboration, are needed. The Pacific Business Group on Health has developed such an effort through its California

Quality Collaborative (CQC). CQC is a healthcare improvement organization dedicated to advancing the quality and efficiency of patient care in California. To date, P4P groups participating in CQC have significantly improved their patient experience results for the physicians that participated in the collaborative.<sup>8</sup> Strategies for spreading the collaborative are under development.

**Figure 8 –Regional Performance Variation (2007) Patient Experience Composite Score**



**KAISER – A LEADER IN PATIENT EXPERIENCE**

**Who**

Kaiser Permanente is a working partnership of two organizations: the not-for-profit Kaiser Foundation Health Plan and Hospitals, and the Permanente Medical Groups — a closed, integrated delivery system with salaried physicians. The northern and southern California medical groups provide care for over 6 million members. Kaiser participates only in the public reporting aspect of the P4P program.

**What**

Kaiser is among the few organizations that score exceptionally well on patient experience while continually improving results. Sixteen Kaiser medical centers in California placed in the top 20% of groups for patient experience in 2007.

**How**

Kaiser understands that patient experience performance is linked to patient satisfaction with access to personal physicians, specialists, and nurse advice lines. Kaiser’s investment in technology provides state-of-the-art options to communicate with members and support self-care, education, and follow up. Patients can manage appointments online, fill/refill prescriptions, and receive lab results via their personal computer. Clinical results include an explanation, comparison to normal ranges, and next steps, if needed. Online health assessment surveys score individual responses, compare results to guidelines, help set patient goals, and provide follow-up progress reports and information. Patients can also email their doctor with non-urgent questions and receive a reply within 24 hours.

Additional success factors include: extensive and ongoing patient surveys; benchmarking; performance feedback to medical centers and individual providers; comprehensive training; and patient services support. Performance results in the patient experience category impact both compensation and the ability to advance within the organization.

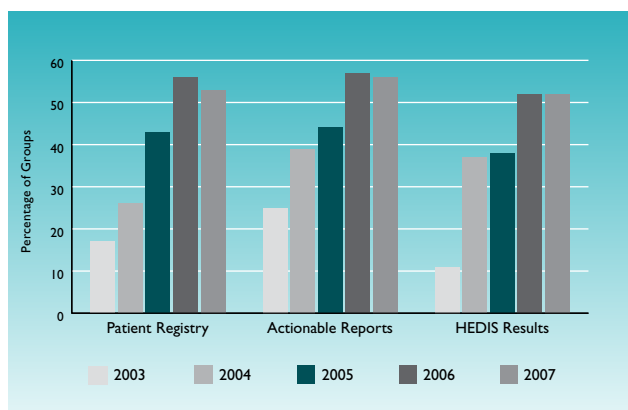
## Information Technology and “Systemness” Results

### Substantial Information Technology Improvements

Nearly two-thirds of physician groups demonstrated some IT capability for measurement year 2007, about double the number of groups in 2003. Additionally, nearly one-third of physician groups have demonstrated robust care management processes. The results in this domain are significant, especially given the correlation between IT adoption and clinical performance. *Adoption of IT systems that support electronic patient registries for chronically ill or at risk patients, and use of electronic decision support systems at the point-of-care are important building blocks for care management processes that offer potential to improve quality of care.*

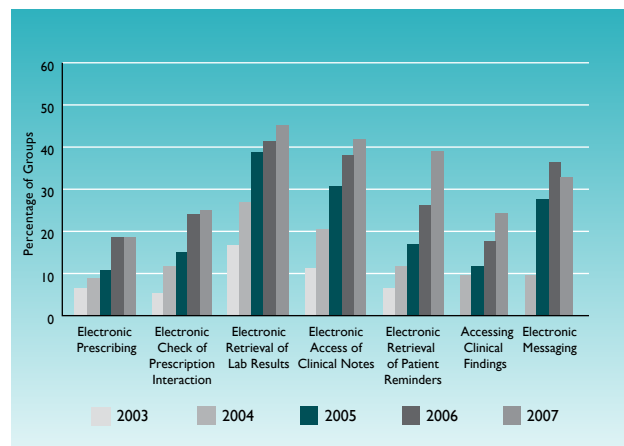
Figure 9 shows a substantial increase in information technology uptake of population management activities over baseline. For measurement year 2007, nearly 54% of groups used electronic patient registries, up from a baseline of 18% for 2003. Group use of online actionable reports has more than doubled: approximately 58% of groups demonstrated use in 2007, versus a baseline of 25% for 2003. Performance rates for population management activities appear to start leveling off in 2007. This may be a result of the movement of the IT measures into the IT systemness domain, with changes in IT measure weighting and scoring.

**Figure 9 – IT Measure: Population Management Activities (2003-2007)**



Most point-of-care activity measures have continued to show substantial gains, as illustrated in Figure 10. Measures include e-prescribing and electronic checks for drug interactions, a practice that significantly improves patient safety. In 2003, only 5% of groups had implemented e-prescribing; in 2006 and 2007, that number increased to 18%, a rate higher than the national average.<sup>9</sup> By comparison, a variety of sources estimate that as of 2007, 12% of office-based doctors nationally were e-prescribing. Other notable performance improvements for point-of-care activities include the ability of groups to retrieve lab results electronically, which more than doubled to 40%, and the ability to access clinical notes electronically, which more than tripled, to approximately 43% of groups.

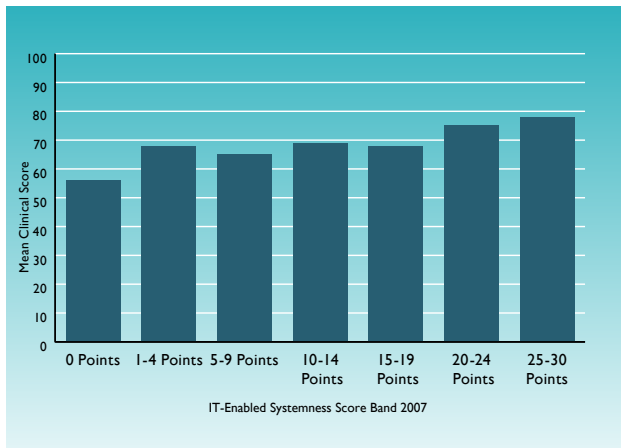
**Figure 10 – IT Measures: Point-of-Care Activities (2003-2007)**



### Correlation between IT Capabilities and Clinical Performance

*P4P groups with advanced IT capabilities outperform groups with no reported IT capability by an average of 22 percentage points on clinical measures, as shown in Figure 11. Physician organizations with IT are better able to capture data, which may be part of the reason their clinical scores are higher. They also have the data needed to monitor their performance effectively, which allows them to identify where outreach is needed to improve performance. Physician groups report having used the P4P incentives to accelerate adoption of IT.*

**Figure 11 – Average Clinical Score by IT-Enabled Systemness Score Band (2007)**



The following case study illustrates the link between IT and improved clinical performance: St. Mary Independent Physicians Association was able to improve diabetes and LDL control substantially through use of electronic disease registries.

## ST. MARY IPA – USING DISEASE REGISTRIES TO CREATE BREAKTHROUGHS IN DIABETES CARE

### Who

St. Mary Independent Physicians Association (IPA) is a small physician group that is managed by Coast Healthcare Management in Lakewood, California.

### What

St. Mary IPA received a P4P award for being the most improved physician group in the Los Angeles region for measurement year 2006. The percentage point improvements in clinical performance results were impressive.

### 2006 – 2007

- HbA1c and LDL screening increased 13%

### 2007 – 2008

- LDL screening increased 16%
- LDL control increased 14%
- HbA1c control increased 19%

### How

St. Mary IPA attributes its improvement in diabetes care to the increase in the use of electronic registries. The use of registries jumped from 5% to 70% of primary care practices within three years. This dramatic increase in the number of physicians using electronic registries helped St. Mary IPA quickly identify both member and primary care physician outliers. Coordination with care management teams steered diabetic patients to group classes, diabetic clinics, one-on-one sessions with a nutritionist, and an endocrine consultant. Registries helped focus practices on population management. P4P incentives for scoring high on quality performance measures motivated reluctant physicians to use the technology.

## IV. THE “PAY” IN PAY FOR PERFORMANCE

P4P incentives come in several forms: financial, public reporting, and public recognition. Financial incentives remain the most influential.

During the early formation of the program, several important principles were adopted. One key principle was agreement that all participating health plans would determine incentive payments based upon aggregate performance of physician groups, not performance for just the patients served by a single health plan. This principle leveraged the power of collaboration, but other aspects of the “pay” in P4P remained unresolved.

From the start of the program, a primary goal was an incentive payment that was large enough to drive breakthrough performance improvement while keeping pace with an expanding measurement set. *However, the incentive amount and how it should be structured and funded remain open issues today.* In 2006, a five year strategic plan was developed, which established important payment-related goals: first, create a safe haven to advance consistent payment methodology, and second, have incentive payments reach 10% of compensation by 2010. A Payment Committee was formed in 2007 after resolving legal concerns over potential antitrust issues, as described later in this section. However, little progress has been made toward achieving the 10% payment target. In fact, total incentive payments declined in 2008.

Payment issues remain a central and ongoing challenge for the program; they will likely require fundamental reform. This is a contentious issue, but participating plans and physicians groups are considering ways to restructure the program.

### Financial Incentive Payment Amounts

Total P4P payment by participating health plans increased annually from 2004 to 2007, but declined in 2008. P4P incentive payments in 2009 are expected to decrease further. The decreased annual payout, combined with the increased number of performance measures in the

measurement set, resulted in the lower financial reward payment per measure shown in Figure 12.

**Figure 12 – Number of P4P Measures and Bonus Payments**

Measurements	2003 <sup>(1)</sup>	2007 <sup>(2)</sup>
Total Number of Measures	25	49
Total Bonus Payout (\$)	\$38M	\$52M
Bonus Payout (\$) Per Measure	\$1.52M	\$1.06M

<sup>(1)</sup> 2003 measurement year bonus paid in 2004

<sup>(2)</sup> 2007 measurement year bonus paid in 2008

Financial incentives paid from 2004 through 2008 (for the prior measurement year) total approximately \$264 million.

### Total Plan Payouts:

\$38 million paid out in 2004

\$54 million paid out in 2005

\$55 million paid out in 2006

\$65 million paid out in 2007

\$52 million paid out in 2008

California physician groups receive most of their compensation for HMO enrollees in the form of professional services capitation. These payments are supplemented by P4P incentive payments, along with incentive payments for non-P4P measures related to medical and administrative efficiency. To put these incentive payments in context, the P4P payouts of \$65 million in 2007, combined with additional non-P4P incentive payouts of \$89 million, create a total of \$154 million in incentive payments for 2007. Total incentive payments equaled about 2% of the total base compensation paid to physician organizations in 2007.

Physician groups received higher or lower percentages of incentive payment based upon performance. Nonetheless, many stakeholders legitimately raise the need for much larger incentive opportunities to reward participating physician groups. Surveys by program evaluators indicate physician groups believe that incentive potential of at least 5% of total compensation is required to motivate significant change.

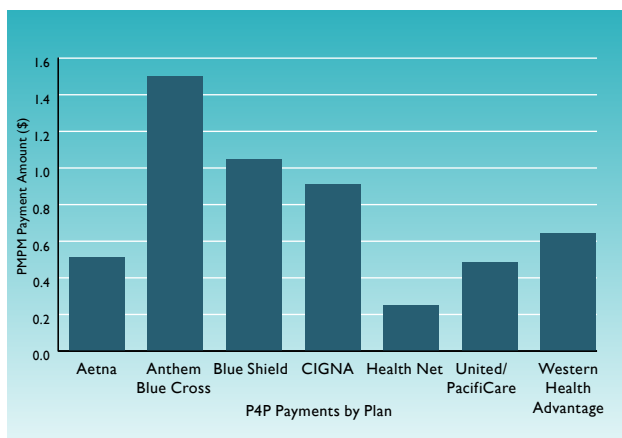
The decrease in the total P4P incentive payment for 2008 was primarily the result of one health plan deciding to shift \$8 million of incentives previously allocated for P4P measures to a non-P4P shared savings program.<sup>10</sup> In shared savings programs, health plans share with physician groups the savings that have been generated by more efficient and appropriate use of resources. This efficiency/gain-sharing approach will be part of the P4P measurement set starting in 2009.

### Variation in Payment Amounts Jeopardizes Program Sustainability

There is a nearly eightfold difference in per member per month payouts between the lowest paying health plan and the highest paying health plan. This inconsistency has created what some have coined a “free rider” issue, with the lower paying plans getting equal benefit from the program at the expense of the higher paying plans.

Lower paying plans have indicated that to justify paying higher P4P incentive amounts, they would need to receive self-reported physician group data to improve HEDIS performance. The highest paying plans have indicated they will likely start paying less if they do not see more equity in plan payments and better HEDIS performance. Decreased payouts jeopardize the viability of the P4P program and must be addressed. Figure 13 illustrates P4P payment disparities between plans in 2008.

**Figure 13 – P4P Payments by Plan (2008 Payout Year)**



### Standardization of Financial Incentive Payment Methodology

Significant variation exists in payment methodology across plans, a source of confusion among participating physician groups. To bring transparency to payment methodologies, IHA and RAND developed an annual retrospective report documenting payment methodologies and amounts by plan. This report, “Transparency Report on Health Plan Payouts,” was first produced in 2004 and is posted, along with subsequent reports, on the IHA website ([www.IHA.org](http://www.IHA.org)).

In 2007, the P4P Steering Committee approved a policy that requires participating health plans to communicate P4P incentive payment budget amounts and methodology in advance of the next measurement year. For example, health plans communicated their incentive payment budget amount and methodology for 2008 (to be reported and paid in 2009) by January 1, 2008. IHA developed a template for plans to submit this information, which is then compiled and posted on the IHA website. Health plans have responded by submitting timely information. However, actual payments have varied somewhat from intended payment amounts, primarily because actual performance results varied from projected results. To address growing requests to standardize payment methodology between health plans, IHA has also convened a Payment Committee.

### Multi-Payer Collaboration and Antitrust Considerations

Early in the program, efforts to move payment discussions forward among plans were stymied by legal considerations related to antitrust laws. Since forming a Payment Committee, IHA has devoted substantial effort to removing these barriers.

Multi-payer P4P collaboration raises serious legal considerations because incentive bonuses are a type of payment. Any discussion or effort to establish uniform payment among competitors can be considered “collusion” to fix price – a per se violation of federal antitrust laws. This restriction initially caused P4P stakeholders not to discuss payment methodology or amount, with

the exception of the most basic recommendations. With the public benefit of P4P at risk, IHA engaged legal counsel, who approached a regional office of the Federal Trade Commission (FTC) on an informal basis for guidance.

In sharing the guidance provided, neither IHA nor its stakeholders offer legal advice. The principle guidelines developed for IHA to assist with this issue are as follows:

- The establishment of a multi-payer Payment Committee was considered allowable, with clear restrictions on what could and what could not be discussed.
- The Payment Committee may make recommendations for P4P incentive payment amounts and methodologies.
- The Payment Committee may make recommendations only. Program participants cannot be required to use these recommendations as a condition of participation, and participation must remain non-mandatory.
- Incentive bonus payments cannot be pooled. Each participating health plan must continue to make payment directly to its contracted physician groups.

### **Payment Methodology Recommendations**

The P4P Steering Committee has approved the following recommendations of the Payment Committee:

#### **Payment for Improvement and Attainment**

Use of an incentive payment calculation methodology developed by the Centers for Medicare and Medicaid Services (CMS). This methodology scores the performance of each physician group in two ways: first, based upon level of attainment, and second, based on the amount of improvement. The higher score is used to determine payment amount. A more detailed explanation of this methodology is included in Appendix C.

#### **Linking Payment to Data Sharing**

A tiered payment structure with higher payments linked to better data sharing by physician groups. Tiered payment will encourage physician groups to share electronically available clinical data with health plans to support health plan HEDIS scores and quality improvement efforts.

#### **Gain-Sharing Methodology for Appropriate Resource Use Measures**

A gain-sharing methodology for appropriate resource use measures (e.g., re-admission rate) has been introduced into the 2009 measure set. This approach is funded by savings generated through increased efficiencies and appropriate resource use. It allows physician groups to share in savings generated and earn additional compensation, without increasing costs to the health plan.

The Payment Committee continues to work on payment methodology and towards a solution to its most vexing problems: the amount and variation in payment between participating health plans.

## V. PUBLIC REPORTING AND PUBLIC RECOGNITION

Financial incentives may be a primary driver for performance improvement, but public reporting and public recognition also influence behavior.<sup>11</sup> A single, annual public report card published online shows pay for performance scores for physician groups by measure and by composite scores for each county. IHA publicly recognizes the top performing and the most improved physician groups each year during the annual P4P Stakeholders meeting and through a press release.

### Public Reporting

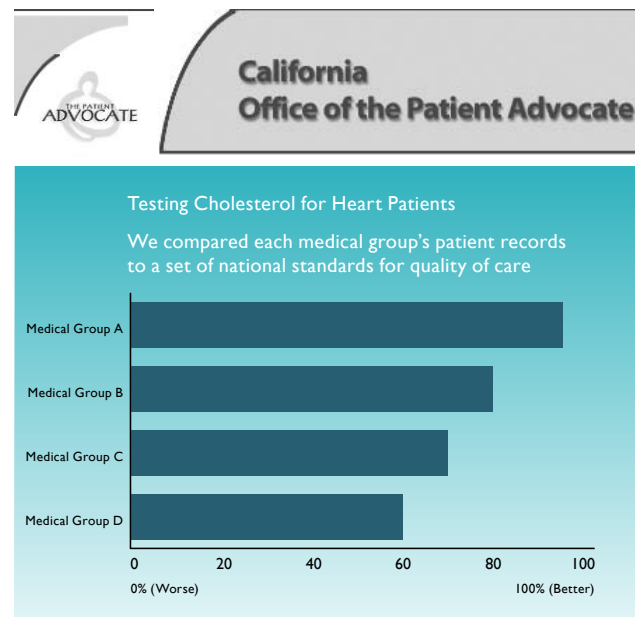
To drive healthcare accountability and transparency while educating consumers, the P4P annual physician group performance results are posted on a public website sponsored by the California Office of the Patient Advocate (OPA) ([www.opa.ca.gov](http://www.opa.ca.gov)). Each year, OPA publicizes the annual report card results for California physician groups. Local media coverage of the performance ratings often calls out the best and worst performers in the area. While public reporting serves as a stimulus for change, anecdotal information suggests that healthcare providers appear to be paying more attention to the report card information than consumers. *Physician groups are intensely interested in seeing how they have performed compared to competitors, as well as how the information appears in the state report card.*

Consumers who visit the OPA website are able to search performance results by county and look at performance for clinical care and patient experience by physician group. Research confirms that consumers are interested in health care information, but it needs to be useful and understandable.<sup>12</sup>

**Figure 14 – OPA Medical Group Ratings**  
([www.opa.ca.gov](http://www.opa.ca.gov))

	Meeting National Standards of Care	Patients Rate Medical Groups
Medical Group A	No report due to incomplete data	Not rated
Medical Group B	★ ★	★ ★
Medical Group C	★ ★ ★	★ ★ ★
Medical Group D	★ ★ ★ ★	★ ★ ★
Medical Group E	★ ★ ★ ★	★ ★ ★
Medical Group F	★ ★ ★ ★	★ ★ ★

**Figure 15 – OPA Individual Measure Ratings**  
([www.opa.ca.gov](http://www.opa.ca.gov))



#### What was measured?

What percentage of medical group adult members had their low density lipid cholesterol (LDL) level checked in the last year after having a heart attack, heart surgery or for their cardiovascular disease? Lower levels of LDL cholesterol are associated with better outcomes in those with heart disease. These results are based on patient administrative records from the medical group or HMO.

#### Why is it important?

High cholesterol damages your blood vessels. Testing your cholesterol regularly can help you from having a heart attack or a stroke. Medical groups that earn high ratings test patient's cholesterol regularly.

## Public Recognition

IHA presents two types of performance recognition awards each year: (1) top performers by measurement domain and overall; and (2) most improved.

Top performers are identified by calculating composite scores in each measurement domain, which are then weighted according to the recommended P4P payment weights. An overall performance composite score for each participating physician group is calculated, and the top 20% of groups with the highest overall composite score are designated as top overall performers statewide.

In 2007, P4P began recognizing physician group improvement to underscore the benefit of program participation beyond just the highest performers. The physician group in each of eight geographic regions that demonstrated the highest relative quality improvement over the previous year was designated an award winner. The award for most improved is named posthumously for Ronald P. Bangasser, MD, a family physician and vice-chairman of the Board of Beaver Medical Group, who also chaired the IHA P4P Technical Quality Committee and served on numerous high profile state and national quality committees. To determine the award winners, an improvement score is calculated for each physician group based on the percent of relative improvement the group achieved in their overall performance composite score from the previous year to the current measurement year.

Public recognition awards for top performing and most improved groups have been well received. Many groups use their awards to acknowledge the hard work of internal team members, and highlight awards in advertising and marketing materials.

## VI. PROGRAM MANAGEMENT AND ORGANIZATION

Multi-stakeholder efforts to measure and reward performance have emerged in a number of regional collaborations in the U.S. In most cases, an existing organization provides the operational base for performance measurement and related quality improvement initiatives. Each regional effort has its own unique focus and character, bringing stakeholder groups together for a common purpose. Such organizations include: Greater Detroit Area Health Council, Indiana Health Information Exchange, Massachusetts Health Quality Partnership, Minnesota Community Measurement, Puget Sound Health Council, and Wisconsin Collaborative for Healthcare Quality.

IHA is often asked, “How do you organize and staff a collaboration of this nature? How is it funded? And what is your budget?” The following is offered to answer these questions.

### Statement of Principles

IHA began the pay for performance program with a statement of vision, goals, core principles, and project objectives. Agreement on these principles was an important step forcing debate over divergent views and competing objectives. It also established ground rules that defined the scope and operation of the program. The initial set of guiding principles was developed in 2002, and later modified in 2006. These principles (Appendix D) continue to maintain program integrity.

### Program Governance

A simple governance process was established at the start of the program. Initially, two committees were created: a Technical Committee and a Steering Committee. The former was responsible for measure set development and specification, data collection, and other technical details of the program. The Technical Committee and staff made recommendations and reported to the Steering Committee, which reviewed the recommendations and was responsible for making final decisions and overseeing the program. The Steering Committee reports to the IHA board, which adopted a “hands-off” policy in respect to day-to-day P4P program

management. The IHA board does, however, periodically review program progress and maintains strategic oversight of the program.

A few changes were made to the P4P committee structure in 2006 - 2007. First, the Technical Committee was re-organized into two separate committees: the Technical Efficiency Committee, responsible for the development of cost efficiency and resource use measures, and the Technical Quality Committee, responsible for the other measure domains. In addition, the size of the committees was reduced to nine voting members each.

Next, an Executive Committee was created as a subset of the Steering Committee to handle long term planning and provide more frequent direction to program staff. In 2007, a Payment Committee was also created to explore and recommend common incentive payment methodologies. Formation of this committee occurred after legal and regulatory issues were resolved, as described earlier in this document.

The P4P governance process is built upon consensus; however, split votes are not uncommon, especially as the program has increased in complexity. The number of committees, subcommittees, and workgroups allows significant and substantive discussion on all issues. However, the formal, multi-level decision-making process, combined with program complexity, affects the ability to act quickly. In addition, the membership on these committees has been weighted toward clinical and technical expertise. This has resulted in some disconnect between the clinical and business side of participating organizations. In retrospect, more workgroups, fewer standing committees, and a streamlined decision-making process might allow more timely development and implementation of program change.

## Staffing and Technical Support

The P4P program was initially staffed by existing IHA personnel and consultants; a full-time program manager was hired in 2005. Currently, the program is supported by a full-time program director and a full-time program manager, with part-time support from other IHA staff members. The program's self-assessment shows that the program has been chronically understaffed, particularly during the initial phases.

Outside consultants handle many aspects of data collection, data aggregation, measure testing and specification, analysis and related activities. The National Committee for Quality Assurance (NCQA) provides primary technical support for clinical quality, IT, systemness and related measures. RAND Corporation and the Pacific Business Group on Health provided technical support for patient experience measures. More recently, technical assistance on the development of efficiency measures has come from Thomson Reuters (formerly Medstat) and independent consultants.

NCQA manages aggregation of clinical results from participating health plans and physician groups. Each physician group or health plan data submission must pass an audit by an NCQA certified auditor. The multiple data sets are aggregated, and resulting scores are presented to stakeholders who have an opportunity for review and comment before scores are finalized.

## Funding

The initial funding for P4P was provided in the form of technical and development grants by the California HealthCare Foundation and a supporting grant from GlaxoSmithKline to cover the expense of committee meetings. As these funds diminished, IHA implemented an "administrative surcharge" against the incentive payment paid by the participating health plans. This surcharge equaled \$.075 per member per year when it was first levied in 2006, and has increased

as grant funding diminished. The current surcharge of \$.345 per member per year totaled \$1.92 million for 2008, and funds the entire program budget.

2008 funding was allocated in the following categories:

EXPENSE TYPE	BUDGET AMOUNT
Personnel expenses	\$ 410,000
Data collection and aggregation	\$ 350,000
Technical support for measure testing and specification	\$ 125,000
Analysis	\$ 25,000
Cost efficiency measure development and testing	\$ 725,000
Committee meetings and governance	\$ 61,000
Office related expenses	\$ 134,000
Legal and accounting	\$ 45,000
Other	\$ 45,000
TOTAL	\$ 1,920,000

## Program Evaluation

A five year evaluation by RAND Corporation and University of California at Berkeley completed at the end of 2008 was funded by the California HealthCare Foundation. The results of this evaluation have been reported in several publications, which are either footnoted or listed at the end of this document.

## VII. LESSONS LEARNED

The California P4P experience has been shared in detail so that others working to measure, report and reward performance may benefit. The following lessons are offered with this view in mind.

### **Physician Groups Respond Favorably to Health Plan Collaboration**

The value of health plan collaboration to create and use a uniform measurement set cannot be overstated. Physician groups consistently report that a uniform measurement set facilitates measurable improvement in an environment of ever-tightening resources.

Equally important is the aggregation of multi-payer data to produce a single “score,” versus individual plan scoring, each using a slice of the physician group’s patient population. This has significantly improved the reliability of the results, and perhaps more importantly, physician trust of the results.

A public scorecard using aggregated results is also viewed favorably by physician groups. It addresses a previous source of contention: multiple report cards each using a portion of the patient population, rendering the results neither reliable nor fair to those being measured.

In combination, these collaborative efforts have improved the relationships of health plans and physician groups in California.

### **Attention to Pay and Metrics Needs to be Balanced**

Early in the program, most activity and resources were targeted to measure testing and development, and less thought was given to the pay side of the equation. To some extent this was driven by antitrust concerns, and unfortunately, disagreements about the appropriate level of incentive were not resolved. Today, we recognize clearly that the pay side of the equation must be balanced with the metrics.

### **Physician Group Engagement Has Strengthened**

According to interviews conducted by program evaluators in 2005 and again in 2007,<sup>13</sup> P4P has inspired California physician groups to

become more highly engaged in quality and data collection. These interviews revealed an increased engagement in the P4P program and quality improvement activities, with 74% responding that the measures are reasonable; widespread support for higher incentive amounts; and increased focus on group quality improvement and IT capabilities. Overall, physician groups gave P4P an average rating of 4 on a scale of 1 to 5, with higher scores indicating a more favorable response.

### **Health Plan Engagement Has Weakened**

Health plan interviews and feedback demonstrate weakening health plan engagement, and point out the need to make fundamental program changes. P4P program strengths identified by health plans include: increased collaboration; a push toward quality improvement; investments in IT; and greater accountability and transparency. Program weaknesses cited include: improvements viewed as marginal; concerns about “teaching to the test;” lack of a positive return on investment; and failure of clinical data sharing to raise plan HEDIS scores. Overall, health plans gave P4P an average rating of 2.5 on a scale of 1 to 5.

Health plan executives have pledged their continued participation in P4P, but emphasized there will not be an increase in incentive payments without a stronger business case. Further, major changes in payment methodology and the addition of efficiency measures are needed to maintain health plan interest in P4P.

### **Electronic Clinical Registries Are a Valuable Tool**

The value of clinical registries to aid population management in chronic care has been argued by Ed Wagner, MD, an architect and advocate of the Chronic Care Model.<sup>14</sup> The California P4P experience suggests electronic registries support better data collection, measurement, internal reporting, patient monitoring, and outreach. The costs are relatively modest, providing a short-term solution that can be integrated with electronic medical records. Evidence for these positive factors is the increase in registry uptake by California P4P physician groups from an 18% usage rate in 2003 to more than 57% today.

### **Measuring Appropriate Use, Overuse and Cost Efficiency is Essential but Difficult**

After almost two years of concerted effort, it became clear the contemplated episode-of-care approach for efficiency measurement and reward was simply not feasible. A uniform set of population-based appropriate resource use measures was quickly developed based on existing measures that individual plans had implemented. Further work is in progress to enhance population-based measures and gain-sharing rewards. For episode-of-care measurement, potential efforts will be redirected toward production of variance and use reports for physician group self-improvement activity, but not for measurement and payment purposes.

### **Geographic Variation May Be Symptomatic of Underlying Provider Payment Disparity**

Clinical quality practice variation, and its apparent association with low socioeconomic areas, has captured the attention of stakeholders. It has inspired more emphasis on payment for

improvement and led IHA to collaborate with researchers at the RAND Corporation and Cornell University to study factors influencing performance differences.

The challenges faced by physician groups dedicated to serving economically challenged service areas are best described by the groups themselves.

### **Clinical and Administrative Data Streams Must Be Merged**

At the program's onset, a decision was made to limit data collection to electronic sources only, excluding any chart review. Chart review on this scale was considered overly expensive. To compensate for the lack of chart review, physician groups were allowed to use electronic clinical data and to self-report results subject to external audit.

Merging plan administrative data and physician group clinical data greatly enhanced performance reporting capabilities. However, self-reported clinical data have yet to be included in the health plan HEDIS data reporting process. As a result,

## **THE MULTICULTURAL IPA – SERVING THE UNDERSERVED**

### **Who**

The Multicultural IPA (MCIPA) serving Central and Southern San Diego highlights a number of problems facing those located in underserved areas. MCIPA is a for-profit IPA with over 6,000 commercial HMO, 1,000 Medicare Advantage and 3,000 MediCal enrollees served by 60 primary care physicians and 80 specialists.

### **What**

In a number of ways MCIPA starts at a disadvantage; small size, low payment rates, a predominantly African American and Hispanic population with higher disease burden and lower socioeconomic status, and a physician shortage. Capacity and financial limitations restrict IT support and staff needed to drive clinical quality improvements measured by P4P. Severe physician shortages mean less access, longer wait times, and lower patient experience results. These challenges limit the opportunity to earn incentives.

"If P4P can find a way to deal equitably with diverse patient populations, it can be a great tool to improve quality and save money. If it doesn't, it may cause more harm than good," according to Dr. Rodney Hood, President of MCIPA.

### **How**

Despite these challenges, MCIPA leadership was able to show its physicians the potential P4P incentive payments that were "left on the table" – approximately \$150,000 per year. Considering this, MCIPA identified opportunities for improvement and is working to help its practices improve performance in P4P. A decision was made to make infrastructure investments, including implementation of an electronic medical record (EMR), in eight practices with significant patient population. Quality improvement staff personnel visit practices implementing the EMR each week to facilitate training and problem resolution. Further, quarterly meetings of primary care physicians explore ways to improve quality performance.

Dr. Hood believes that P4P must adapt for physician groups in medically underserved areas to deal with the barriers they face: paying for improvement is not enough. At the same time he sees P4P as part of payment reform, part of a larger effort needed to improve the quality of care and outcomes for all populations.

the full advantage of P4P improvements are not being captured in plan HEDIS results. This limits the plans' ability to demonstrate improved HEDIS scores and undermines health plan perception about the return on investment for incentive payments. Limitations on clinical data sharing also mean plans must continue to spend resources conducting chart audits to support HEDIS measurement.

### **System-Wide Performance Breakthrough Remains Elusive**

P4P in California has not yet achieved a primary objective – breakthrough quality improvement. New P4P measure domains are creating incentives that encourage systems of care intended to transcend the claim that P4P promotes “teaching to the test.” Examples include the IT Systemness and Coordinated Care domains.

A common view is also emerging among P4P stakeholders. While structuring incentives to foster care re-design at the group level is important, achieving program breakthroughs will require:

- Higher bonus payments.
- Targeted interventions in regions of the state with lower overall performance.
- Quality improvement support to parallel measurement and reward.

### **Pay For Performance is Not the Answer, But is an Important Step Toward a Solution**

The attention that has been given to pay for performance is often criticized as an effort to promote a “silver bullet” directed at gaps in quality. Instead, the enthusiasm for P4P reflects the most important lesson learned: that building a valid, reliable foundation for performance measurement forms the basis of trust necessary to engage physicians and other healthcare providers in quality improvement. While the rate of improvement has not met expectations, it underscores the primary tenet of quality improvement: “You cannot improve what you don’t measure.” P4P has provided the crucial framework necessary for making quality improvement a reality.

## VIII. LOOKING AHEAD

The California P4P program, along with other regional collaboratives, faces significant uncertainty in this time of economic downturn and potential health reform. The California program has matured and must now reinvent itself. Our stakeholders look to maintain the positive benefits of collaboration while integrating new ideas to achieve elusive breakthrough improvements. The foundation of collaboration and trust already established will serve as the platform for change as P4P undergoes fundamental redesign.

### **In the future:**

- Cost and quality must be integrated to address affordability and the new economy.
- Gain-sharing from improved efficiencies will be the main funding source for future financial incentives.
- Performance measurement and reward will migrate from reporting and “bonus” payments into performance-based contracting and payment.
- Efforts to improve and merge electronic administrative and clinical data sources will continue in conjunction with regional efforts toward interoperable data exchange, yielding gradual improvements.
- Outcomes and specialty physician services will be incorporated into performance measurement as clinical data collection methods mature and new measures are developed and tested.
- Initiatives to address low performers must be developed to supplement P4P performance measurement and rewards.

P4P is not the final destination, but rather an important step along the path to accountability, continuous quality improvement, and payment reform.

## IX. APPENDICES

### Appendix A – IHA Pay for Performance Committees

#### **P4P EXECUTIVE COMMITTEE**

Steve McDermott (Chair)  
Hill Physicians Medical Group

Alain Enthoven, Ph.D.  
Stanford University

Alan Glaseroff, M.D.  
Humboldt-Del Norte IPA

Arnold Milstein, M.D.  
Mercer Health & Benefits

David Hopkins, Ph.D.  
Pacific Business Group on Health

David Joyner  
Blue Shield of California

Jerry Penso, M.D.  
Sharp Rees-Stealy

Michael Belman, M.D.  
Anthem Blue Cross

Sam Ho, M.D.  
UnitedHealthcare

#### **P4P STEERING COMMITTEE**

Arnold Milstein, M.D. (Chair)  
Mercer Health & Benefits

Michael Belman, M.D.  
Anthem Blue Cross

Daniel Bluestone, M.D.  
Sante Community Physicians

Michael-Anne Browne, M.D.  
Blue Shield of California

Thomas J. Davies  
Corporate Healthcare Management

Alain Enthoven, Ph.D.  
Stanford University

Marge Ginsburg  
Center for Healthcare Decisions

Alan Glaseroff, M.D.  
Humboldt-Del Norte IPA

Christine Griger, M.D.  
Palo Alto Medical Foundation

Steven Halpern, M.D.  
CIGNA Healthcare of California, Inc.

Sam Ho, M.D.  
UnitedHealthcare

David Hopkins, Ph.D.  
Pacific Business Group on Health

Don Hufford, M.D.  
Western Health Advantage

Michael Kern, M.D.  
John Muir Physician Network

Philip Madvig, M.D.  
The Permanente Medical Group

Robert Margolis, M.D.  
HealthCare Partners

Steve McDermott  
Hill Physicians Medical Group

Charles Payton, M.D.  
Health Net of California

Les Schlaegel  
Stanford University

Julie Wade  
GlaxoSmithKline

Melissa Welch, M.D.  
Aetna, Inc.

Tom Williams  
Integrated Healthcare Association

#### **Liaison Members**

Sandra Perez  
Office of the Patient Advocate

Chris Perrone  
California HealthCare Foundation

#### **Consultants / Staff**

Jennifer Benjamin  
National Committee for Quality Assurance

Cheryl Damberg, Ph.D.  
RAND

Keely Macaulay  
National Committee for Quality Assurance

Jas Nihalani  
Integrated Healthcare Association

Kristiana Raube, Ph.D.  
University of California, Berkeley

Dolores Yanagihara  
Integrated Healthcare Association

**TECHNICAL QUALITY COMMITTEE**

Alan Glaseroff, M.D. (Chair)  
Humboldt-Del Norte IPA

Ellen B. Fagan  
CIGNA Healthcare of California, Inc.

David Hopkins, Ph.D.  
Pacific Business Group on Health

Joel Hyatt, M.D.  
So. California Permanente Medical Group

Stuart Levine, M.D.  
HealthCare Partners

Ranyan Lu, Ph.D.  
UnitedHealthcare

Eileen O'Connor  
Health Net of California

Ernest Valente, Ph.D.  
Blue Shield of California

Ann Woo, Pharm.D.  
Hill Physicians Medical Group

**Non –Voting**

Cheryl Damberg, Ph.D.  
RAND

John Ford, M.D.  
Practicing Physician

Paul Solari, M.D.  
Genentech

**Consultants / Staff**

Jennifer Benjamin  
National Committee for Quality Assurance

Keely Macaulay  
National Committee for Quality Assurance

Jas Nihalani  
Integrated Healthcare Association

Dolores Yanagihara  
Integrated Healthcare Association

**TECHNICAL EFFICIENCY COMMITTEE**

David Hopkins, Ph.D. (Chair)  
Pacific Business Group on Health

Daniel Bluestone, M.D.  
Santé Community Physicians

Michael-Anne Browne, M.D.  
Blue Shield of California

Becky Kapp  
Anthem Blue Cross

Paul Katz  
Intelligent Healthcare

Michael van Duren, M.D.  
Sutter Connect

Melissa Welch, M.D.  
Aetna, Inc.

**Consultants / Staff**

Sean Kennedy  
Thomson Reuters

Mahil Senathirajah  
Thomson Reuters

J. William Thomas, Ph.D.

Dolores Yanagihara  
Integrated Healthcare Association

**PAYMENT COMMITTEE**

Alain Enthoven, Ph.D. (Chair)  
Stanford University

Steve McDermott (Ex-Officio)  
Hill Physicians Medical Group

Bart S. Asner, M.D.  
Monarch Healthcare

Timothy Duffey  
Merck & Company, Inc.

David Joyner  
Blue Shield of California

Jeff Kamil, M.D.  
Anthem Blue Cross

Arnold Milstein, M.D.  
Mercer Health & Benefits

Bart Wald, M.D.  
HealthCare Partners

**Consultants / Staff**

Tom Williams  
Integrated Healthcare Association

Dolores Yanagihara  
Integrated Healthcare Association

## Appendix B - 2009 P4P Measurement Set

### YEAR 7 MEASURES: 2009 MEASUREMENT YEAR / 2010 REPORTING YEAR

#### CLINICAL DOMAIN

---

Clinical PO Encounter Threshold for reporting<sup>1</sup>:

4.0 Encounters per member per year  
(using Encounter Rate by Service Type specs)

Weighting: 40%

1. Childhood immunization status with 24-month continuous enrollment
2. Appropriate treatment for children with upper respiratory infection
3. Breast cancer screening
4. Chlamydia screening
5. Use of appropriate medication for people with asthma
6. Cholesterol management: LDL screening for patients with cardiovascular conditions
7. Cholesterol management: LDL control <100 for patients with cardiovascular conditions
8. Colorectal cancer screening
9. Appropriate testing for children with pharyngitis
10. Avoidance of antibiotic treatment for adults with acute bronchitis
11. Use of imaging studies for low back pain
12. Medication monitoring (ACE/ARBs, digoxin, diuretics)
13. Evidence-based cervical cancer screening

#### PATIENT EXPERIENCE DOMAIN

---

Weighting: 20%

1. Getting appointment with a specialist
2. Rating of specialist
3. Timely care and service composite
4. Doctor-patient Interaction composite
5. Care coordination composite
6. Rating of PCP
7. Rating of all healthcare
8. Office staff composite
9. Health promotion composite

#### IT-ENABLED SYSTEMNESS DOMAIN

---

Weighting: 20%

1. Data integration for population management
  - a. Reporting based on electronic information
  - b. Identifying important conditions
2. Electronic clinical decision support at the point-of-care
3. Care management
  - a. Coordination with practitioners
  - b. Chronic care management
  - c. Continuity of care
4. Physician measurement and reporting

## YEAR 7 MEASURES: 2009 MEASUREMENT YEAR / 2010 REPORTING YEAR

### COORDINATED DIABETES CARE DOMAIN

---

Weighting: 20%

1. HbA1c screening
2. HbA1c poor control (>9)
3. HbA1c control (<8)
4. LDL screening
5. LDL control <100
6. Nephropathy monitoring
7. Diabetes registry and related activities
  - a. Blood pressure in registry
  - b. Actionable reports to MDs
  - c. Individual MD measurement
8. Diabetes care management program

### APPROPRIATE RESOURCE USE DOMAIN (“Efficiency” Domain)

---

Weighting: Standard gain-sharing arrangement in development

1. Inpatient utilization—acute care discharges
2. Inpatient utilization—bed days
3. Outpatient surgeries utilization—% done in ASC
4. Emergency department visits
5. Inpatient readmissions within 30 days
6. Generic prescribing

### REPORTABLE NON-PAYMENT MEASURES

---

Measures to be collected and publicly reported, but not recommended for payment

- I. Medicare measures:
  - a. Breast cancer screening
  - b. Diabetes care HbA1c screening
  - c. Diabetes care HbA1c poor control
  - d. Cholesterol management LDL screening (patients with cardiovascular conditions and/or diabetes)
  - e. Cholesterol management: LDL Control <100 (patients with cardiovascular conditions and/or diabetes)
  - f. Nephropathy monitoring for patients with diabetes
  - g. Colorectal cancer screening

### TESTING MEASURES

---

Measures to be collected for testing and analysis.

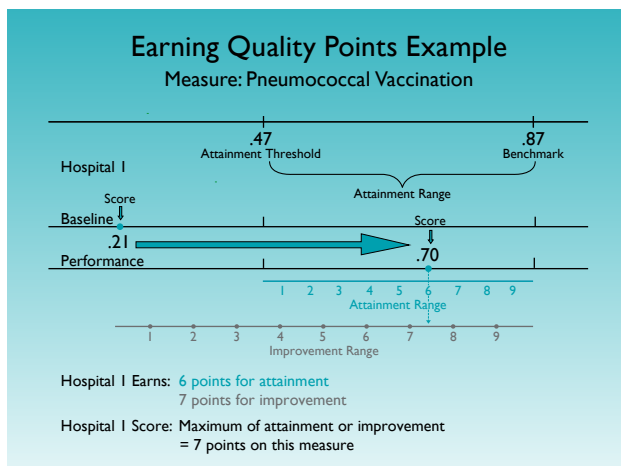
To be determined.

I PO Encounter Threshold refers to the number of encounters PMPY required to be included in data aggregation and public reporting. For the purposes of payment, individual health plans may use a different encounter threshold. Additional information can be found in the P4P Manual at [www.iha.org](http://www.iha.org)

## Appendix C – Value Based Purchasing: Payment Methodology for Improvement and Attainment

Value-based purchasing (VBP), which links payment to performance, is a key policy mechanism that the Centers for Medicare and Medicaid Services (CMS) is proposing to transform Medicare from a passive payer of claims to an active purchaser of care. The Congress, through the Deficit Reduction Act of 2005, Section 5001(b), authorized the Secretary of Health and Human Services to develop this method to commence in Fiscal Year 2009 for Medicare hospital services provided by hospitals paid under the Inpatient Prospective Payment System.

VBP evaluates a hospital's performance on each measure based on the higher of an "attainment score" in the measurement period or an "improvement score" determined by comparing the hospital's current measure score with its prior-period baseline performance. This approach would encourage a broad range of hospitals to engage in value-based purchasing – even those that begin with a low absolute level of performance. The scores on individual measures across different performance domains (clinical process-of-care, patient perspectives of care, and 30-day mortality outcomes at VBP Program start) are combined to compute a hospital's VBP Total Performance Score, which is then used to determine the percentage of the VBP incentive payment earned by the hospital.



Source: U.S. Department of Health and Human Services Report to Congress: Plan to Implement a Medicare Hospital Value-Based Purchasing Program, November 21, 2007.

- **Benchmark:** the reference point defining high level of performance
- **Attainment threshold:** the minimum level of performance required to receive attainment points
- **Attainment range:** the scale between the attainment threshold and benchmark
- **Improvement range:** the scale between the hospital's prior year score (baseline) on the measure and the benchmark

## Appendix D – P4P Vision, Goal, Core Principles, Program Objectives

### Vision

The achievement of breakthrough improvement in healthcare performance.

### Central Goal

The overall goal of Pay for Performance (P4P) is to significantly improve physician group performance in quality of health care and patient experience through public recognition and financial reward.

### Core Principles

#### Collaboration

P4P is accomplished through purchasers, health plans, physician groups and consumers working together.

#### Measurement

The measurement set is comprehensive and dynamic, including measures of clinical quality, patient experience and infrastructure to support patient care. Continuous evaluation will ensure alignment, relevance and effectiveness, raising the bar on performance over time.

#### Reward

Health plans will offer financial incentives tied to performance results. The financial incentives will be significant and sustained to promote performance driven organizations and justify investment in system re-engineering.

### **Accountability**

All stakeholders have a role:

- Purchasers will promote health plan participation in P4P.
- Physician organizations will implement appropriate internal performance measurement systems, including individual physician measures.
- A public scorecard will report physician group performance for consumers and providers in making informed choices.

### **Program Objectives**

Strategic selection criteria – include measures that are:

- Clinically relevant
- Affect a significant number of people
- Scientifically sound and tested before implementation
- Feasible to collect using administrative data
- Physician groups and health plans can impact
- Capable of showing improvement over time
- Are important to California consumers
- Aligned with national measures (where feasible)

### **System Reform**

Encourage system re-engineering over incremental improvement. Move from an individual disease management approach to cost-cutting measures. Reward better outcomes, customer service, structure, and efficiency, for greater change and consumer relevance. Apply risk adjustment so that payment reflects population mix and rewards better performance with patients who require special care.

### **Consumer-relevant**

Add customer service and other credible measures that evaluate better service to members, administrative efficiency, and quality-related utilization.

### **Predictability and stability**

Ensure predictability and stability in the measurement set. Phase-in multiple part measures, moving from process to outcomes as appropriate. Leave each measure in the set for at least 3 years. Evaluate annually to adjust based on experience, including weighting and specifications. Consider testing for one year when measures have not been used before.

### **Standardize for comparison**

Provide the greatest comparability between physician groups and enhance consumer and provider benefit by having participating health plans use a standardized measurement set.

### **Align what matters**

Work to align Pay for Performance (P4P) measures among plans, providers, and purchasers with the measures required by accreditation, Health Employer Data Information System (HEDIS), public and private purchasers and regulators, including movement to individual physician level performance.

## REFERENCES

- <sup>1</sup> Baker G, Delbanco S. Pay for Performance: National Perspective 2006; Longitudinal Survey Results with 2007 Market Updates. San Francisco, CA: MedVantage, 2007.
- <sup>2</sup> Rosenthal MB, Landon BE, Normand SI, et al. Pay for Performance in Commercial HMOs. *New England Journal of Medicine* 2006; 355(18):1895-902.
- <sup>3</sup> Regan M, Randall-Lewis R. Collaboration: The Power of Data Aggregation. *HealthLeaders Media*, April 22, 2008.
- <sup>4</sup> Damberg CL, Raube K, Teleki S, dela Cruz E. Taking Stock of Pay for Performance: A Candid Assessment from the Front Lines. *Health Affairs*, March 2009.
- <sup>5</sup> Rosenthal MB, Landon BE, Normand SI, et al. Pay for Performance in Commercial HMOs. *New England Journal of Medicine* 2006; 355(18):1895-902.
- <sup>6</sup> Damberg CL, Raube K, Teleki S, dela Cruz E. Taking Stock of Pay for Performance: A Candid Assessment from the Front Lines. *Health Affairs*, March 2009.
- <sup>7</sup> National Quality Forum. National Voluntary Consensus Standards for Ambulatory Care Part I & II, 2008.
- <sup>8</sup> Diane Stewart of Pacific Business Group on Health. Improving Patient Experience Ratings. *Presentation to IHA P4P Steering Committee*, October 2008.
- <sup>9</sup> Landro L. *Incentives Push More Doctors to E-Prescribe*. The Wall Street Journal.com, January 21, 2009.
- <sup>10</sup> Integrated Healthcare Association. Transparency Report on 2008 Health Plan Payouts. [www.IHA.org](http://www.IHA.org), March 2009
- <sup>11</sup> Fung CH, Lim YW, Mattke S, Damberg C, and Shekelle PG. The Impact on Quality of Publicly-Released Performance Data: A Systematic Review of the Peer-Reviewed Literature. *Annals of Internal Medicine*, 2008;148:111-123
- <sup>12</sup> Robert Wood Johnson Foundation. How to Engage the Public's Interest in Health Care Quality Data: Financial Incentives. Report of Grant Results, October 2008.
- <sup>13</sup> Damberg CL, Raube K, Teleki S, dela Cruz E. Taking Stock of Pay for Performance: A Candid Assessment from the Front Lines. *Health Affairs*, March 2009.
- <sup>14</sup> Wagner E. Simple Steps to a Patient Registry: Ticket to Care Coordination, Quality Reporting and Pay for Performance. *Healthcare Intelligence Network*, October 2008.

### Additional Publications on Pay for Performance:

Damberg CL, Raube K, Williams T, and Shortell SM. Paying for Performance: Implementing a Statewide Project in California. *Quality Management in Health Care*, (14:2) Spring 2005.

Williams T, Raube K, Damberg CL, et al. Pay for Performance: Its Influence on the Use of Information Technology in Physician Organizations. *Journal of Medical Practice Management* Mar/April 2006.

Teleki SS, Damberg CL, Pham C, Berry SH. Will Financial Incentives Stimulate Quality Improvement? Reactions from Frontline Physicians. *American Journal of Medical Quality*, Vol. 21, No. 6, Nov/Dec 2006.

Wheeler JRC, White B, Rauscher S, Nahra TA, Reiter K, Curtin K, Damberg CL. Pay for Performance as a Method to Establish the Business Case for Quality. *Journal of Healthcare Financing*. 2007; 33(4):17-31.



Integrated Healthcare Association  
300 Lakeside Drive, Suite 1975  
Oakland, CA 94612

*office:* 510.208.1740  
*fax:* 510.444.5842  
*website:* [www.iha.org](http://www.iha.org)