Q Manage Health Care
Vol. 14, No. 2, pp. 66–79

Paying for Performance: Implementing a Statewide Project in California

Cheryl L. Damberg, PhD; Kristiana Raube, PhD; Tom Williams, MBA, MPH; Stephen M. Shortell, PhD

The US health care system falls far short of providing care consistent with national standards of care and available knowledge. There continue to be wide variations in how care is provided, and a host of studies and reports point to substantial deficits in the quality of care being delivered.[1–3] The study by McGlynn and colleagues[4] that examined the delivery of care in 12 communities across the United States found that patients receive recommended care only about 50% of the time, irrespective of whether the care was preventive, acute, or chronic. These deficits were found across medical conditions as well. Moreover, work by Casalino et al with a national population of Physician Organizations (POs) found that they used less than 50% of recommended care management processes (eg, disease registries, guidelines, automated reminders), with great variability across organizations and chronic illness conditions.[5]

In recognition of these deficits, the Institute of Medicine released a landmark document *Crossing the Quality Chasm*,[6] which was a national call to action. This report calls for significant change at all levels of the health delivery system, with particular emphasis on system redesign that will drive substantial improvements to close the quality gap. Of note, the report called for creating and aligning incentives for quality and increasing the transparency of quality information for quality improvement, accountability, and consumer choice.

In 2002, the Robert Wood Johnson Foundation (RWJ) and the California HealthCare Foundation (CHCF) funded 7 demonstration projects under the Rewarding Results program to implement and evaluate financial and nonfinancial incentives for quality.[7] The funding for this initiative totals more than $8.8 million and is impacting the care of 22 million Americans. Among the 7 projects, the largest is the Integrated Healthcare Association's (IHA) Pay for Performance (P4P) program, which currently covers over 6.5 million or close to one quarter of all Californians. The sheer scope of P4P gives it great weight within the most populous state in the nation. The 7 participating health plans—Aetna, Blue Cross, Blue Shield, Cigna, Health Net, PacifiCare, and Western Health Advantage—have contractual relationships with all of the major capitated physician organizations, touch 45,000 physicians, and for some Independent Practice Associations (IPAs) represents 100% of their capitated revenue.

This article describes the implementation of the IHA P4P program and explores the difficult decisions and collaborative structures that were created to make statewide P4P a reality in California. In contrast to several of the other Rewarding Results P4P demonstrations that involve only one health plan (eg, Excellus Health Plan in Rochester, New York, or Blue Cross Blue Shield of Michigan), this project is unique in that it involves multiple, competing commercial health plans in a large statewide initiative, thus representing the kind and scale of interorganizational coordination that may be needed to have substantial impact.

PREVIOUS STUDIES

Several articles have explored the incentives implicit in differing payment structures, while others have suggested a research agenda to further examine the impact of financial incentives on physician behavior, specifically quality.[8,9] Others have

From RAND, Santa Monica, Calif (Dr Damberg); University of California, Berkeley (Drs Raube and Shortell); and the Integrated Healthcare Association (Mr Williams).

Corresponding author: Cheryl L. Damberg, PhD, Senior Policy Researcher, RAND, 1776 Main St, Santa Monica, CA 90407 (e-mail: damberg@rand.org).

reviewed the myriad of physician incentive programs to reward quality.[10] The published literature on physician incentive programs speaks primarily to the impact of financial incentives on utilization and productivity.[12–18] Few rigorously designed studies exist that examine incentive programs directed at physicians to improve the quality of care delivered. Of the few published studies examining the effect of incentives for quality, the incentives were limited in scope (eg, childhood immunizations, cancer screening) and/or the results were mixed.[19–23] To date, there have been no empirical studies of the impact of a physician incentive program across a diverse patient and payer population. Nor have there been publications that describe the implementation hurdles that must be overcome.

Physician organizations in California were more likely to be exposed to external incentives to improve quality than those in the rest of the country, even before the start of the P4P program.[24] Nonetheless, differential payments to California medical groups during the 1990s focused more on utilization management and less on the quality of care.

PAY FOR PERFORMANCE PROJECT

The IHA Pay for Performance (P4P) project is a collaborative effort among 7 California health plans, 215 POs, purchasers, and consumer advocates to implement financial and nonfinancial incentives to improve quality. The impetus for P4P in the California market was a collectively defined need to create the business case for quality, with the view that financial incentives would drive performance improvements and investments in information technology (IT) to support quality management and improvement. While the individual health plans could implement their own programs to improve quality, these disparate efforts would result in large data collection burdens on POs, due to lack of coordination. Moreover, no single health plan had enough market share within any single PO to effect a behavior change. The P4P project was an effort to reduce confusion in the marketplace, both for providers and consumers, that occurred as a result of competing health plans producing individual PO scorecards with different measures and conflicting results.[25]

The concept of a collaborative effort among California health plans and POs was conceived in 2000 by members of the IHA, a multi-stakeholder leadership group. This followed an effort the previous year by the purchaser community, led by Pacific Business Group on Health (PBGH), to pay a 2% bonus to POs based on quality performance. IHA is a not-for-profit association of health plans, hospital systems, POs, and related health care companies, policymakers, and representatives of employers, consumers, and government. The role of the IHA is to engage the various sectors of health care delivery in dialogue and collective action with the mission of advancing the quality of health care delivered to its constituents in California.

IHA historically has acted as a convener for the health care industry in California on a variety of policy and health care delivery issues including the uninsured, technology assessment, and risk adjustment. The concept of an industry-coordinated effort to create a quality incentive program evolved from an annual IHA leadership summit and was advanced at a subsequent session sponsored by an IHA member organization. The agreement to collaborate on a uniform set of quality measures occurred in 2001 and a governance structure for the program was organized.

The program was given the name Pay for Performance (P4P) and a Steering Committee was formed that includes leaders from participating health plans (Aetna, Blue Cross of California, Blue Shield of California, CIGNA, HealthNet, and PacifiCare), POs, government, purchasers and consumer groups. The committee is assisted in its work by representatives from organizations with expertise in quality measurement and reporting including PBGH, The National Committee on Quality Assurance (NCQA), Centers for Disease Control, RAND, and University of California – Berkeley. The Executive Director of IHA was designated as the chair of this committee with responsibility to organize quarterly meetings and provide staff support.

The P4P Steering Committee was charged with establishing strategic direction for the P4P program,

providing governance and setting policy as necessary. In response, the Steering Committee decided to create a program predicated on the belief that the group would be more powerful than the single plan, that common quality metrics could drive performance, and that significant financial incentives would attract and maintain the attention of POs. They determined that the overall goal of the P4P was to significantly improve PO performance in the delivery of quality health care through public recognition and financial reward, thereby establishing the project's core objectives:

- *Collaboration*: P4P is accomplished through purchasers, health plans, POs, and consumers working together.
- *Measurement*: The measurement set will be comprehensive and dynamic and include measures of clinical quality, patient experience, and infrastructure to support patient care. Continual evaluation will assure 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 reengineering.
- *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 comparative physician group performance for use by consumers.

In addition to the Steering Committee, a Technical Committee was created to govern development of a uniform performance measurement set, standards for data collection, and other technical components of the P4P program. The membership of this committee includes multi-stakeholder representation drawing on the collective technical expertise of organizations participating in the program. This Committee has been supported by the Provider Group Oversight (P-GO) team, which is composed of individuals with measurement expertise from NCQA and the PBGH. The P-GO team has been an important resource for the implementation of the P4P, as they conduct many of the required analyses to inform project decisions. Involvement of staff from the NCQA provides an important link to NCQA's work with the Health Employer Data Information Set (HEDIS) measures used for health plan accreditation. The Technical Committee developed the initial performance measurement set, which was submitted for public comment and eventually approved by the Steering Committee late in 2001. At this point, 6 California health plans had agreed to participate by implementing a quality incentive program based either partially or entirely on the uniform P4P measurement set. As of 2004, a seventh health plan, Western Health Advantage, joined the P4P.

Consistent with the program objectives above, the P4P includes both financial (payments by the health plans to the POs) and nonfinancial incentives (a report card that provides public recognition). Under the P4P, POs receive bonus payments from the contracted health plans on the basis of a set of measures of clinical performance, patient experience, and IT investment. The performance measures were chosen on the basis of several criteria:

- *Strategic focus*:
  - Align with national measures (where feasible)
  - Clinically relevant
  - Affect a significant number of people
  - Scientifically sound
  - Feasible to collect using administrative data
  - POs and health plans can influence
  - Capable of showing improvement over time
  - Important to California consumers
- *System reform*: Encourage system reengineering over incremental improvement. Move from an individual disease management approach to crosscutting measures and 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 doing better with patients that require special care.
- *Consumer-relevant*: Add customer service and other credible measures that better evaluate service to members, administrative efficiency, and quality-related utilization.
- *Predictability and stability*: Assure predictability and stability in the measurement set. Phase-in multiple part measures—moving from process toward outcomes as appropriate. Leave each measure in the set for at least 3 years. Evaluate annually to adjust on the basis of experience, including weighting and specifications. Consider testing for 1 year where measures have not been used before.
- *Standardize*: Provide the greatest comparability between POs and enhance consumer and provider benefit by having participating health plans use a standardized measurement set.
- *Alignment*: Work to better align P4P measures among plans, providers and purchasers with the measures required by accreditation, HEDIS, and public and private purchasers and regulators, including movement to individual physician-level performance.

In keeping with these criteria, the first year of P4P experience was assessed using 6 clinical measures (Table 1). All 6 measures use the HEDIS specifications, are produced only using administrative data, and include patients who are enrolled on a continuous basis. The patient experience measures come from the California Consumer Assessment Survey (CAS), which is derived from the national group-level CAHPS survey. Physician organizations that want to receive bonus payments from the health plans for the consumer experience metrics need to participate in the annual CAS survey fielded by the California Cooperative Healthcare Reporting Initiative (CCHRI). Finally, based on the belief that good quality is a result of the ability to manage care and identify those patients at risk for chronic disease or in need of preventive services, there are several information technology (IT) measures on which POs are asked to self-report. For future measurement years, P4P is adding more clinical measures, is making the IT measures more stringent, and creating an additional bonus opportunity for POs that provide performance data to individual physicians. For the 2004 measurement year the weights associated with the performance domains shifted, with slightly less weight on clinical measures and more on IT capability.

The P4P began in 2002 with a testing year, in which many decisions were made regarding performance indicators. This was followed by the first measurement year in 2003. In August 2004, health plans and POs received 2003 performance data aggregated for all of the participating plans. In fall 2004, a public report card was published by the California Office of Patient Advocate (OPA) on the basis of the Year 1 (2003) performance results. The report card used the aggregated data across the 6 health plans to develop ratings for the clinical performance of 194 of the 215 POs participating in the P4P, and it included data from the CAS. These results were posted on a public Web site (http://www.opa.ca.gov/report_card/) to support consumers in their health plan and PO selection.

YEAR 1 DATA SUBMISSION AND RESULTS (2003)

Data submission

The NCQA fulfilled the role of an independent "data aggregator" for the program by collecting 2003 measurement year results for all components of the clinical, patient experience and IT domains of the P4P measurement set. Clinical performance measurement results for the 2003 measurement year were submitted to the NCQA by both health plans and self-reporting POs by the end of May 2004. Clinical data were submitted for 215 POs representing 100% of the enrolled P4P population. The patient experience data were taken from CAS for 133 POs, which represent 89.5% of the enrolled population. Finally, the IT data was self-reported to the NCQA by 100 POs, representing 79.2% of the enrolled population (Table 2).

Information in the clinical and IT domains was submitted to the NCQA subsequent to review and approval by outside, independent auditors. The

Table 1

IHA PAY FOR PERFORMANCE MEASUREMENT SET

	2003 Measurement/ 2004 reporting year	**2004 Measurement/ 2005 reporting year**	**2005 Measurement/ 2006 reporting year**
Clinical performance	1. Childhood immunizations with 24-month continuous enrollment 2. Cervical cancer screening 3. Breast cancer screening 4. Asthma management 5. HbA1c screening 6. LDL screening Encounter threshold ≥ 2.7 encounter PMPY*	1. Childhood immunizations with 24-month continuous enrollment 2. Cervical cancer screening 3. Breast cancer screening 4. Asthma management 5. HbA1c screening and control 6. LDL screening and control < 130 7. Chlamydia screening Encounter threshold ≥ 3.25 encounter PMPY*	1. Childhood immunizations with 24-month continuous enrollment 2. Cervical cancer screening 3. Breast cancer screening 4. Asthma management 5. HbA1c screening and control 6. LDL (includes diabetics) screening and control < 130 7. Chlamydia screening 8. Appropriate treatment for children with upper respiratory infection
Weighting	50%	40%	50%
Patient experience	1. Specialty care 2. Timely access to care 3. Doctor-patient communication 4. Overall ratings of care	1. Specialty care 2. Timely access to care 3. Doctor-patient communication 4. Overall ratings of care	1. Specialty care 2. Timely access to care 3. Doctor-patient communication 4. Care coordination 5. Overall ratings of care
Weighting	40%	40%	30%
Information technology investment	1. Integrate clinical electronic data sets at group level for population management 2. Support clinical decision making at point of care through electronic tools Requires 2 activities with at least 1 in each measurement domain; each activity is worth 5%	1. Integrate clinical electronic data sets at group level for population management 2. Support clinical decision making at point of care through electronic tools Requires 4 activities of which at least 2 are in measurement domain 2; each activity is worth 5%	1. Integrate clinical electronic data sets at group level for population management 2. Support clinical decision making at point of care through electronic tools Requires 4 activities of which at least 2 are in measurement domain 2; each activity is worth 5%
Weighting	10%	20%	20%

(*Continues*)

Table 1

IHA PAY FOR PERFORMANCE MEASUREMENT SET (*continued*)

	2003 Measurement/ 2004 reporting year	**2004 Measurement/ 2005 reporting year**	**2005 Measurement/ 2006 reporting year**
Bonus opportunity			Physician measurement bonus: POs can receive up to an additional 10% for a demonstrable program to both measure and provide regular feedback on clinical metrics and patient experience to individual physicians.
Testing Measures—Measures to be collected but not reported		1. Appropriate treatment for children with upper respiratory infection (HEDIS specification) 2. Antidepressant medication management (HEDIS specification) 3. Physician incentive bonus: Provider groups can receive a 10% add-on if an individual physician financial incentive program is in place for either clinical performance or patient experience	1. Flu shots for those aged 50 to 64—Collect sample size information through 2005 CAS survey 2. Colorectal Screening—Collect sample size information through 2005 CAS survey 3. Nephropathy monitoring for diabetics—Use HEDIS Administrative specification
Development Measures—Measures for which specifications will be developed			1. Antidepressant medication management (alternative to HEDIS) 2. Obesity 3. Diabetic retinal exam—modified from HEDIS administrative specification

*PMPY indicates per member per year, POs, physician organizations; CAS, California Consumer Assessment Survey; HEDIS, Health Employer Data Information Set.

Table 2

DATA SUBMISSION SUMMARY (2003)

	Number of groups	Percentage of groups	Percentage of enrollment
Clinical measures	215	100	100
Patient experience*	133	62	89.50
Information technology investment	100	47	79.20

*Some CAS survey results were combined to allow a consistent unit of analysis across measure domains.

information received in each domain was then integrated into a consolidated data set, which was provided to the participating health plans and POs in August, 2004.

2003 Clinical performance results (50%)

To be eligible for credit for a clinical performance measure, a PO needed a minimum of 30 patients and an average of 2.7 encounters per patient, specific to that measure. Performance information was collected for a total of 215 POs for the clinical measure set and 194 had sufficient sample size and patient encounters for reportable results in at least one clinical measure. Of the groups with reportable results, 74 scored at a significantly higher level for at least 4 of the clinical measures, as compared to the mean performance score for all groups. The greatest variations in clinical scores were for the diabetes HBA1c screening, childhood immunizations, and cervical cancer screening. The smallest variations in scores across POs were for asthma care and breast cancer screening. Clinical performance was measured by administrative data only; therefore, some of the variation in clinical performance may reflect relative capabilities of the participating POs to collect and report administrative data (Table 3).

2003 Patient experience results (40%)

POs were required to participate in the CAS to receive credit for results in this performance domain. Of the 215 POs that reported clinical results in year one, 133 were scored in this domain. Although there was not much variation in the performance scores among POs in the area of patient experience with care, POs continue to perform well below optimal levels, especially in the area of access to primary and specialty care, where 25% to 35% of patients reported having problems. These results demonstrate that there is still significant room for POs to improve. Consistent with the previous year's experience on the CAS, POs in Northern California tended to perform better, on average, than groups located in Southern California. The factors that are associated with the regional variation in performance have not yet been determined (Table 4).

Table 3

CLINICAL PERFORMANCE YEAR 1 (2003)

Measure	No. of groups	Mean	SD	90th percentile	75th percentile	10th percentile
Asthma: All ages	145	66.66	6.86	75.13	71.21	58.59
Diabetes care: HbA1c screening	184	65.78	19.66	82.97	77.19	38.33
Cholesterol management: LDL screening	53	67.66	15.12	79.55	78.05	53.33
Breast cancer screening	183	64.38	10.18	75.75	71.52	50.91
Cervical cancer screening	185	62.41	15.02	78.60	73.11	42.86
Childhood immunizations: MMR	148	73.08	13.53	87.43	82.43	51.00
Childhood immunizations: VZV	148	69.02	13.86	84.18	80.24	48.08

Table 4

PATIENT EXPERIENCE YEAR 1 (2003)

Name	No. of groups	Mean	SD	90th percentile	75th percentile	10th percentile
Communication with doctor	133	85.58	3.79	89.66	87.98	80.17
Rating of doctor	131	80.03	5.09	86.20	83.07	72.95
Rating of health care	133	69.98	6.30	78.07	73.98	65.09
No problem seeing specialist	131	59.46	6.91	66.75	63.64	50.03
Rating of specialist	126	70.98	5.90	78.59	75.33	63.20
Timely care and service	133	69.53	5.69	62.47	73.68	76.24

Data from Consumer Assessment Survey, 2004.

2003 Information technology investment results (10%)

For the 2003 measurement year, POs demonstrating any one of the qualifying information technology (IT) activities received half credit, and those achieving 2 activities received full credit for IT. A total of 100 POs of the 215 POs submitted results for the IT domain, with 67 groups receiving full credit (full 10% of bonus), 7 receiving half credit (5% of bonus), and 26 receiving no credit (0%). The implementation of actionable reports utilizing integrated clinical data was the most common activity reported for groups (51%) in the data integration category. Retrieving lab results (36%) and accessing clinical notes (36%) electronically were the 2 most common activities, for POs receiving credit for point-of-care technology.

Incentive payment results

Each individual participating health plan developed its own formula to determine incentive payment amounts and eligibility (Table 5). Payments for the first measurement year were calculated by 4 plans using the aggregated clinical plan results from all 6 participating plans. One plan continued its previous practice of quarterly payments using its own plan experience, and another plan paid annual bonuses using its plan experience to meet contractual obligations, before the collective data set was available.

Table 5

INDIVIDUAL HEALTH PLAN BONUS PROGRAMS (AS OF SEPTEMBER 2003)

Health plan	Use of P4P common measures	Other measures in bonus plan	Data source for bonus	Maximum bonus potential	Payment threshold or percentiles
Aetna	Yes	No	Aggregated data	Up to 3.5% of capitation rate	Percentile ranking
Blue Cross	Some, but not all	Yes	Blue Cross data	$4.50 PMPM*	Percentile ranking
Blue Shield	Yes	No	Aggregated data	$2.00 PMPM	Percentile ranking
CIGNA	Yes	Yes	Aggregated data and CIGNA data	$1.60 PMPM minimum for the top performing groups	Threshold
Health Net	Yes	No	Aggregated data	$2.25 PMPM	Threshold
PacifiCare	Some, but not all	Yes	Multiple data sources	$21 million pool to be allocated	Threshold

*PMPM indicates per member per month.

Total incentive payment amounts for the 6 participating plans against the P4P specific quality measures are estimated at about $40 million for the 2003 measurement year, and total health plan payments for all measures of quality are estimated at over $100 million for 2003.

LESSONS FROM IMPLEMENTATION OF P4P

There are many lessons that can be drawn from the initial implementation of the IHA P4P project. The large number of competing organizations that have collaborated in this initiative may be unprecedented in health care. The lessons from this real-time experiment are instructive to others contemplating a large-scale intervention to create a business case for quality.

Lesson 1: Importance of a credible and neutral convenor

At the heart of the process is the "credible convenor" role: an organization with the power to convene major stakeholders because it has professional standing and because it is trusted not to favor any one stakeholder's agenda over the others. The IHA, an organization that itself represents diverse stakeholders, created the role of a neutral convenor for itself in 1995, and it is a key reason why the different stakeholders have remained committed to the P4P development process.

Lesson 2: Reconciling different philosophies and perspectives

One of the most difficult decisions for the various stakeholders was one of philosophy: would quality improve faster if high goals versus more inclusive goals were set? Some committee members, whom we will call the *Darwinians*, believe that the bar should be set at as high a level as possible to aggressively drive performance improvement. The Darwinians also believe that thresholds for performance should be made progressively more challenging over time. The Darwinians acknowledge that this approach may force out of business some late adopters, those who are less organized, or smaller POs that are more likely to have poorer performance if patient enrollment migrates from poorer- to better-performing groups. However, the Darwinians believe that this ultimately may result in better quality of care if the poorer-performing entities are disbanded or absorbed by better-performing POs. They argue that larger, well-organized groups that embrace quality management and which tend to have higher performance scores should be rewarded for their commitment to quality.

In contrast, other stakeholders, whom we will call the *Social Democrats*, assert that "a rising tide lifts all boats." The Social Democrats believe that broad participation is important, especially at the outset, particularly among small, low-performing groups, and consequently, they want to structure the program so that it provides incentives for all groups to improve care delivery. One of their fundamental beliefs is that not only should absolute performance be rewarded, but also improvement, regardless of the level of the base score. Their goal is to improve quality throughout the state, not simply to consolidate the high-performing POs. The discussions of the P4P program reflect a mixture of both perspectives.

Differing viewpoints about implementing the project have also arisen with the partnership between P4P and the California OPA, a state agency whose goal is to inform and educate consumers about their rights and responsibilities as HMO enrollees. Established in 2000, the OPA publishes an annual Quality of Care Report Card to provide consumers comparative information on the performance of California's largest HMOs and, more recently, medical groups that form the provider networks of the HMOs. To avoid confusion among consumers over potentially competing report cards with different displays of the performance information, the P4P and the OPA agreed to collaborate to produce a single consumer report card. The OPA's report card is distributed

through its Web site, in printed versions, as well as through local pharmacies, community-based organizations, and public libraries. The OPA's legislative responsibility is to advocate for the consumer and so must independently decide on the content and structure of the report card. It reserves the right to make final decisions regarding the report card format and content. The independence of this entity in displaying the results of the P4P program, outside the governance structure of the P4P program generated concerns among some stakeholders over the loss of control in the presentation of the results to consumers, although the end result satisfied all stakeholders.

Lesson 3: Allowing differentiation

Although the health plans are collaborating with respect to the performance metrics to ensure uniformity for a core set of measures, antitrust regulations preclude plans from collaborating on how they pay the POs. As a result, there are 6 different formulas that determine the amount of payment that POs receive under the P4P (Table 5). For example, in the first year Aetna paid the bonus as a percentage of the capitation rate, whereas Blue Cross paid up to a set amount per member per month. In some cases, payment was allocated after specific thresholds were met: CIGNA provided payments only to those POs performing above the 50th percentile. In contrast, Blue Shield paid varying amounts at particular thresholds; for example, performance at the 30th percentile earns 25% of the potential payment for a specific clinical measure.

In addition to differences in payment approaches, P4P did not require all health plans to use only the P4P metrics when determining the basis on which POs were to be rewarded. Several of the health plans included additional metrics they used to calculate bonus payments, although it is important to note that some of these additional metrics derive from legacy quality incentive programs in place prior to P4P. The extent to which plans included additional measures varies, with the most extreme example being Blue Cross, in which 72.5% of the potential bonus for the 2003 performance was tied to non-P4P measures.

Lesson 4: Testing the measure specifications

One of the most important roles in the P4P project has been played by the P-GO team. This small group of analysts is funded by a grant from the California Healthcare Foundation to assist the P4P with developing measures specifications, testing new measures, designing the process for data submission, integration and verification, and producing an integrated "report card" back to plans and POs. These tasks are critical elements of the design of any P4P program, and it requires personnel and financial resources to support this work, which is the backbone of program implementation. Without their input, the P4P would have relied solely on HEDIS or other specifications and would not have understood whether those measures could be implemented directly at the PO level. The P-GO team tested the feasibility and usability of the 6 clinical measure specifications derived from HEDIS. They worked with 6 POs and the 6 participating health plans, which agreed to test the specifications for the 6 clinical measures during the pilot period using 2002 data. In particular, the P4P wanted to understand whether the PO populations were large enough to support producing stable estimates of clinical performance for each clinical measures and whether the POs could follow the specifications within their own data environment (to allow POs the opportunity to capture more encounters on which to base performance scores). This pilot process allowed the team a chance to understand the technical support needs for the POs as they implemented measures specifications as well as the cost trade-offs of implementing the specifications at the PO and plan levels. As a result of their work, not only were the clinical measures specified in a manner that allowed smooth, feasible data collection, but multiple technical assistance sessions were provided around the state to the POs, which maximized PO opportunities to self-report clinical measures.

Lesson 5: Ensuring the integrity of data used to score Physician Organizations

One of the challenges in basing quality payments on performance scores derived from administrative data is the capture and transfer of those data between the PO and the health plan and ultimately to the data aggregator (Fig 1). In a capitated environment where encounter data are not the basis of payment (as occurs in a fee-for-service environment), encounter data capture by health plans was less than complete for some POs. Some POs had invested in state-of-the art data systems and believed their data were more reliable and complete than those reported by the health plan. This belief was confirmed by prior experience some California POs had with some health plans that had produced clinical score cards using administrative data. During these discussions, the plans and POs agreed there were problems with the "identification of encounters," and "hand off" and "processing" of encounter data from POs to plans, which likely accounted for the low number of encounters shown in health plan files. As a result, the P4P Steering Committee agreed that POs be allowed the option to report 1 or more of the 6 clinical measures themselves, with the provision that these data be audited. If a PO was unable or unwilling to report the clinical measures, the health plan data would be used. The higher score, regardless of source, would be reported and used as the basis for payment. Of the 215 POs, 56% self-reported at least 1 clinical measure. In total, 26% of all reportable clinical rates came from group-supplied data.

To ensure that the administrative data capture was relatively complete, the health plans set a minimum threshold for POs to participate in the P4P. The threshold was set at 2.7 encounters per member per year (PMPY) that the plans would receive from the POs. An analysis by the P-GO team found that 80% of the groups would qualify for public reporting if a 2.7 PMPY threshold were set. Moving forward, to promote improvement in data capture, the P4P has increased the encounter data threshold to 3.25 encounters PMPY upon which clinical performance scores are computed. This step to encourage more complete submission of encounter data is also expected to move all plans toward use of the integrated plan results rather than relying on their own plan results for computing payments.

Lesson 6: Need for an independent data aggregator

Early in the formation of the P4P project, it became clear that an independent, third-party entity would be required to aggregate data at the PO level across multiple health plans. Since the P4P project includes data provided by competing organizations, an external third party was essential to the process. The data aggregator's role is to create an annual report for each participating PO and health plan that includes integrated scores for the clinical measures, patient experience, and care management (Fig 1). The data aggregator provides each health plan with the PO's benchmark (ie, percentile ranking) and threshold scores so that it can make its payment decisions. Four of the 6 health plans relied only on the aggregated data to make first-year bonus payments, while 2 used their own health plan data (Table 5). Each PO received its own specific results, as well as the norms for all groups for comparative information. Finally, PO scores were provided for the public report card.

Lesson 7: Need for good communication among all parties

Not surprisingly, an undertaking of this magnitude requires extensive and ongoing communication, not only of the decisions, but of the processes used to determine the rules of the program, in large part to assure the "buy in" of those involved. In addition to the internal meetings of the P4P Steering and Technical Committees (and their associated work groups), the IHA holds an annual public stakeholders meeting each fall to update POs on the P4P project and to solicit input and feedback. Feedback is also sought through the IHA's public call for comment on performance measures for future years of the program. The IHA has also deployed representatives of its Committees to serve as public speakers at local events, meetings and conferences, such as the annual meeting

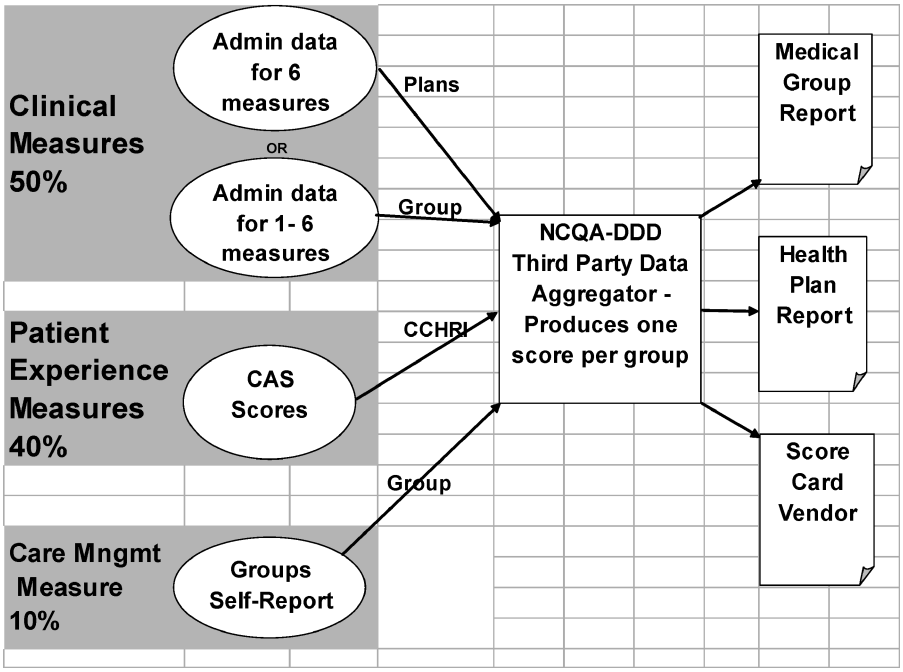

Figure 1. Data collection overview—2003 Measurement year/2004 Payout year.

of the California Association of Physician Groups, to engage the provider community, provide for dialogue, and to serve the broader purposes of ongoing education.

FUTURE ISSUES

A variety of issues confront the program as it evolves from an early developmental stage to a more established standing. These issues underscore complexities and opportunities not addressed at the programs' onset, but present important considerations for future program development.

Rewarding Improvement: An earlier discussion described the differing schools of thought that favor rewarding relative versus improved performance. A consensus has developed among stakeholders supporting the importance of rewarding improvement over time to motivate initial low and middle performers. To accomplish this, agreement must be reached on how to measure improvement, for example whether or not to establish a minimum threshold first and then reward improvement after the threshold is attained.

Pace and Scope of Measure Development: Purchasers and health plans advocate for the addition of many new measures, while the POs favor a more deliberate approach. Establishing the appropriate pace and scope of the performance measure set requires balancing the expectations of these stakeholders in conjunction with the administrative capacity of the program.

Measurement Set Weighting: Determining the proper relative weighting for the performance measure set domains will be dictated by changing priorities. The importance of incentives to promote investment in IT is a topic open to considerable debate.

Limitations of Administrative Data: Capturing accurate laboratory and pharmacy data is essential to move from process-oriented measures to outcome measures in the clinical domain. The use of administrative data only also limits the potential to develop, or add, certain new measures, including BMI screening or depression management.

Expansion to Other Patient Populations: The current program is limited to commercial HMO and POs members, and requests to expand the program to Medicare, Medicaid, PPO and self-insured populations have been rejected to date. These decisions likely will be revisited in coming years.

Centers for Medicare and Medicaid Services (CMS) Involvement: The importance of CMS and its growing interest in quality incentive payments highlights the need to align measures and rewards in the future.

Including Specialist Physicians: The California P4P program currently emphasizes primary care with a limited focus on specialists. Expansion of the measurement set is essential to engage specialists in quality improvement efforts.

The issues facing California IHA P4P stakeholders present an outline of outstanding issues that must be ultimately considered by P4P initiatives nationwide. As an early harbinger of the P4P movement, California offers the opportunity to bring to the surface and explore these important issues in a project of significant scale.

CONCLUSION

In summary, the IHA's P4P initiative is the largest and most complex of any program in the United States that is paying physician organizations for delivering better quality of care. There were several key factors that have led to the successful implementation of this program:

- A credible convener (ie, IHA) to bring competing groups together
- Reconciling different philosophies and perspectives regarding whether to be more or less inclusive while promoting higher standards for quality
- Allowances for some differentiation in health plan incentive programs
- Resources to test performance measures in advance of their going "live"
- Permitting data submission from both POs and health plans
- A neutral third party that could aggregate the data from the health plans and POs
- Ongoing communication among all stakeholder groups

IHA's Pay for Performance program successfully met its first-year goals: paying $40 million in incentive payments in 2004 and laying the foundation for future measurement and rewarding of improved performance.

REFERENCES

1. Schuster MA, McGlynn EA, Brook RH. How good is the quality of health care in the United States? *Milbank Q.* 1998;76(4):517-563.
2. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending, I: The content, quality and accessibility of care. *Ann Intern Med.* 2003;138(4):273–287.
3. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending, II: Health outcomes and satisfaction with care. *Ann Intern Med.* 2003;138(4):288–299.
4. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *N Engl J Med.* 2003;348(26):2635–2645.
5. Casalino L, Gillies RR, Shortell SM, et al. External incentives, information technology, and organized processes to improve health care quality for patients with chronic diseases. *JAMA.* 2003;289(4):434–441.
6. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy of Sciences; 2001.
7. Rewarding Results. Available from www.leapfroggroup.org.
8. Robinson JC. Theory and practice in the design of physician payment incentives. *Milbank Q.* 2001;79(2):149–77.
9. Conrad DA, Christianson JB. Penetrating the "black box": financial incentives for enhancing the quality of physician services. *Med Care Res Rev.* 2004;61(suppl, pt 3):37S–68S.
10. Rosenthal MB, Fernandopulle R, Ryu Song H, Landon B. Paying for Quality: Providers' Incentives for quality improvement. *Health Affairs.* 2004;23(2):127–141.
11. Epstein AM, Lee TH, Hamel MB. Paying physicians for high-quality care. *N Engl J Med.* 2004;350(4):406–410.
12. Hillman AL, Pauly MV, Kerstein JJ. How do financial incentives affect physicians' clinical decisions and the financial performance of health maintenance organizations? *N Engl J Med.* 1989;321(2):86–92.
13. Hemenway D, Killen A, Cashman SB, Parks CL, Bicknell WJ. Physicians' responses to financial incentives. Evidence from a for-profit ambulatory care center. *N Engl J Med.* 1990;322(15):1059–1063.

14. Hellinger FJ. The impact of financial incentives on physician behavior in managed care plans: a review of the evidence. *Med Care Res Rev.* 1996;53(3):294–314.
15. Conrad DA, Maynard C, Cheadle A, et al. Primary care physician compensation method in medical groups: does it influence the use and cost of health services for enrollees in managed care organizations? *JAMA.* 1998;279(11):853–858.
16. Gosden T, Pedersen L, Torgerson D. How should we pay doctors? A systematic review of salary payments and their effect on doctor behaviour. *Q J Med.* 1999;92(1):47–55.
17. Gosden T, Forland F, Kristiansen I, et al. Impact of payment method on behaviour of primary care physicians: a systematic review. *J Health Serv Res Policy.* 2001;6(1):44–55.
18. Stearns SC, Wolfe BL, Kindig DA. Physician responses to fee-for-service and capitation payment. *Inquiry.* 1992;29(4):416–425.
19. Hillman AL, Ripley K, Goldfarb N, Nuamah I, Weiner J, Lusk E. Physician financial incentives and feedback: failure to increase cancer screening in Medicaid managed care. *Am J Public Health.* 1998;88(11):1699–1701.
20. Fairbrother G, Hanson KL, Friedman S, Butts GC. The impact of physician bonuses, enhanced fees, and feedback on childhood immunization coverage rates. *Am J Public Health.* 1999;89(2):171–175.
21. Fairbrother G, Friedman S, Hanson KL, Butts GC. Effect of the vaccines for children program on inner-city neighborhood physicians. *Arch Pediatr Adolesc Med.* 1997;151(12):1229–1235.
22. Hillman AL, Ripley K, Goldfarb N, Weiner J, Nuamah I, Lusk E. The use of physician financial incentives and feedback to improve pediatric preventive care in Medicaid managed care. *Pediatrics* 1999;104(4):931–935.
23. Fairbrother G, Siegel MJ, Friedman S, Kory PD, Butts GC. Impact of financial incentives on documented immunization rates in the inner city: results of a randomized controlled trial. *Ambul Pediatr.* 2001;1(4):206–212.
24. Gillies RR, Shortell S, Casalino L, Robinson JC, Rundall T. How different is California? A comparison of U.S. physician organizations. *Health Affairs* 2003. Available at: http://content.healthaffairs.org/cgi/content/full/hlthaff.w3.492v1/DC1.
25. Payne Simon L, Monroe AF, "California provider group report cards: what do they tell us?" *Am J Med Qual.* 2001;16(2):2001:61–70.