

---

# Quality and Outcomes

---

## Patient Outcomes and Evidence-Based Medicine in a Preferred Provider Organization Setting: A Six-Year Evaluation of a Physician Pay-for-Performance Program

*Amanda S. Gilmore, Yingxu Zhao, Ning Kang, Kira L. Ryskina, Antonio P. Legorreta, Deborah A. Taira, and Richard S. Chung*

---

**Objective.** To determine whether health plan members who saw physicians participating in a quality-based incentive program in a preferred provider organization (PPO) setting received recommended care over time compared with patients who saw physicians who did not participate in the incentive program, as per 11 evidence-based quality indicators.

**Data Sources/Study Setting.** Administrative claims data for PPO members of a large nonprofit health plan in Hawaii collected over a 6-year period after the program was first implemented.

**Study Design.** An observational study allowing for multiple member records within and across years. Levels of recommended care received by members who visited physicians who did or did not participate in a quality incentive program were compared, after controlling for other member characteristics and the member's total number of annual office visits.

**Data Collection.** Data for all PPO enrollees eligible for at least one of the 11 quality indicators in at least 1 year were collected.

**Principal Findings.** We found a consistent, positive association between having seen only program-participating providers and receiving recommended care for all 6 years with odds ratios ranging from 1.06 to 1.27 (95 percent confidence interval: 1.03–1.08, 1.09–1.40).

**Conclusions.** Physician reimbursement models built upon evidence-based quality of care metrics may positively affect whether or not a patient receives high quality, recommended care.

**Key Words.** Physician incentive plans, quality indicators, evidence-based medicine

---

Within the past decade, several Institute of Medicine (IOM) reports have recommended quality-based incentive programs as effective tools to improve quality of care (Kohn, Corrigan, and Donaldson 2000; IOM 2001; Corrigan, Eden, and Smith 2002). In response, many health plans, as well as the Centers for Medicare and Medicaid Services (CMS), have turned their focus to measures that evaluate physician performance in various aspects of care quality such as patient satisfaction and processes of care (CMS 2003; Casalino et al. 2003; Webber 2005). While the programs have produced an abundance of anecdotal evidence suggesting that performance-based reimbursement can affect physician behavior (Morrow, Gooding, and Clark 1995; Fairbrother et al. 1999; Forsberg, Axelsson, and Arnetz 2001; Amundson et al. 2003; Roski et al. 2003), the majority of the programs were implemented in a health maintenance organization (HMO) setting (Levin-Scherz, DeVita, and Timbie 2005; Rosenthal et al. 2005), where evaluation of physician adherence to clinical guidelines is easier to make (Sommers and Wholey 2003), compared with a preferred provider organization (PPO) setting, where responsibility for patient care is more likely to be shared between multiple physicians because patients have freedom to see providers without referral. In addition, HMO plans are more likely to have systematic interventions to improve care processes and outcomes, such as reminders, benefit coverage for screenings, and disease management programs (Casalino et al. 2003), which may contribute to increased adherence to guidelines (Carlisle et al. 1992; Merrill et al. 1999). On the other hand, compared with capitated provider reimbursement in HMO plans, reimbursement arrangements in PPO plans may be better aligned with incentive programs aimed at improving compliance with guidelines that encourage appropriate utilization of certain services (e.g., follow-up, laboratory test, etc.), where base compensation and incentive bonus share common direction (Dudley 2005).

Because PPO structure poses multiple challenges to measurement of provider service quality, the effectiveness of incentive programs on quality of care in a PPO setting has been less rigorously studied despite the increasing role of PPOs as the plan of choice in the United States (Hellinger 1998). This study expands upon the current literature by evaluating the quality of care

---

Address correspondence to Antonio P. Legorreta, M.D., M.P.H., Department of Health Services, School of Public Health, University of California, Campus Box #951772, Los Angeles, CA 90095. Amanda S. Gilmore, M.P.H., Yingxu Zhao, Ph.D., Ning Kang, M.S., and Kira L. Ryskina, B.A., are with Health Benchmarks Inc., Woodland Hills, CA. Richard S. Chung, M.D., and Deborah A. Taira, Sc.D., are with the Hawaii Medical Services Association, Honolulu, Hawaii.

provided by physicians who participated in a novel PPO-based quality incentive program executed by a large nonprofit health plan in Hawaii. We examined whether the quality of care received by patients who visited only physicians who participated in the incentive program improved over a 6-year period compared with patients who visited only physicians who did not participate in the program.

## METHODS

### *Background: The Incentive Program*

Hawaii Medical Services Association (HMSA)—Blue Cross Blue Shield of Hawaii—is the largest provider of health care coverage in Hawaii, ensuring approximately half the state's population, with more than 65 percent of HMSA members electing PPO coverage. HMSA PPO network includes approximately 95 percent of physicians in Hawaii.

In 1998, HMSA launched a physician incentive program to encourage the delivery of high-quality and cost-effective medical care to the health plan's PPO members. General characteristics of the Physician Quality and Service Recognition (PQSR) program have been previously described by Chung et al. (2004). Information about the PQSR program was widely disseminated to Hawaii physicians and professional societies via mass mailing and town and specialty group meetings to encourage participation. Participation in the PQSR program was voluntary (physicians had the option to sign up for this program or not, with the understanding that they would be scored relative to their peers and may qualify for a bonus payment based on performance), and physician enrollment was renewed on an annual basis. Each year, physicians with an active PPO contract who were newly eligible to participate and those who chose not to participate in the previous year received an invitation letter from HMSA.

Physician participation in the incentive program increased considerably since the program's inception: from 50.4 percent of eligible providers in 1998 to 77.7 percent participating in 2003 (see Table 2). All major physician specialties were eligible for participation in the PQSR program. As with overall physician participation in the PQSR program, program participation for most specialty subgroups increased over the study period, with an average increase of approximately 28 percentage points between years 1998 and 2003. In the first year after program implementation (1998), the same proportion of generalists and specialists participated in the program (30 and 29 percent,

respectively). By 2003, a significantly greater proportion of specialists compared with generalists participated in the program (59 versus 50 percent,  $p < .0001$ ).

To participate in the program, physicians were required to fill out an agreement enrollment form that acknowledges the voluntary nature of enrollment and the conditions of award payment. Physicians also received a manual of program operations that included indicator specifications and other logistic information with each year's final report. The PQSR program included four components: clinical performance metrics (board certification, overall member satisfaction rating, and performance on a set of quality indicators measurable by administrative claims data), patient satisfaction (member rating of access to care, communication, and medical services provided by physician), business operations (use of electronic records and participation in multiple HMSA health plans), and practice patterns (medical and drug utilization). In an annual provider satisfaction survey, physicians scored the clinical quality indicators, which are the focus of this study, a 3.9 on a scale of 1–5 in terms of their influence on practice behavior (a rating of 5 indicated the highest level of influence).

The quality indicators examined here are specialty-specific, evidence based, and adapted from existing and widely accepted national clinical guidelines and other available scientific and peer-reviewed literature. Several of the indicators overlap with the Health Plan Employer Data and Information Set (HEDIS<sup>®</sup>) indicators (NCQA 1999–2001), as well as with the priority areas identified by IOM as crucial in improving health care quality in its 2003 report (Adams and Corrigan 2003). The indicators were clearly defined with strict inclusion and exclusion specifications to target the appropriate patient population using administrative claims data.

Because members enrolled in a PPO plan were not assigned a primary care provider by the health plan and were free to seek care from specialists without a referral, the PQSR program adopted a team-oriented approach to appropriately attribute the receipt of recommended care by the member to physicians. For each quality indicator, all physicians of applicable specialty who had recorded a medical encounter with the member were evaluated for the medical care delivered to the member in that measurement period. In other words, PPO physicians' patient populations typically overlapped with other physicians; hence, several physicians could receive credit for care delivered to one member. Only physicians of applicable specialties with 15 or more denominator patients were scored for each indicator.

Eleven of the first 12 quality indicators originally implemented as part of the PQSR program were chosen for evaluation because data were available

for all 6 years; because the indicators applied to the widest denominator populations; and because they were clinically meaningful. These indicators measured: (1) breast cancer screening, (2) cervical cancer screening, (3) hemoglobin A1c (HbA1c) testing for members with diabetes, (4) colorectal cancer screening, (5) continuity of supply of antihypertensive drugs, (6) continuity of supply of lipid lowering drugs, (7) retinal exam for members with diabetes, (8) childhood varicella-zoster virus (VZV) immunizations, (9) childhood measles, mumps, and rubella (MMR) immunizations, (10) use of angiotensin-converting enzyme (ACE) inhibitor in congestive heart failure (CHF), and (11) use of long-term control drugs for asthma. Criteria for each indicator are further described in Table 1. A 12th indicator, measuring complications after cataract surgery, was excluded from the study because the vast majority of plan members who qualified for this measure visited program-participating physicians, thus making a meaningful comparison between program-participating and nonparticipating physicians unfeasible.

Physicians who chose to participate in the PQSR program were compensated with a direct financial reward that was calculated by ranking each practitioner's overall score on all program components relative to the scores of other participating practitioners that fall into the same attainable score range. Based on the percentile ranking among all physicians, the monetary reward ranged from 1 to 5 percent of the physician's base professional fees in 1998–2001, and was increased to 7.5 percent in 2002–2003. The average payment to participating physicians increased from \$2,428 to \$4,785 over the 6-year period. This change reflected an increase in the program budget and the addition of a bonus payment for improvement in performance rate starting in 2001. General characteristics of the PQSR program for 1998–2003 are summarized in Table 2.

Along with the financial incentive, participating physicians received a detailed report including performance rate, percentile rank relative to other participating physicians of applicable specialties, and the median rate for each quality indicator. The report also included feedback on physician performance in patient satisfaction, business operations, and practice pattern components of the PQSR program. Physicians could also request the names of their patients who accounted for the numerator (members who received recommended care) and denominator (members who were eligible to receive recommended care) in each quality indicator computed for the individual physician.

Wide acceptance of the program by physicians and high enrollment rates were achieved through a focused effort on the part of HMSA to maintain

Table 1: Quality Indicators Used in the PQSR Program

<i>Indicator</i>	<i>Denominator</i>	<i>Numerator</i>
Breast cancer screening	Continuously enrolled women aged 50–69 with exception of members with a bilateral mastectomy	Members in the denominator who received at least one screening mammogram during the reporting period
Cervical cancer screening	Continuously enrolled women aged 18–65 with exception of members with hysterectomy with no residual cervix	Members in the denominator who had at least one Papanicolaou smear during the reporting period
Colorectal cancer screening	Continuously enrolled members aged 52–57 with exception of members with any primary diagnosis of colorectal cancer at any point in member history	Members in the denominator who had at least one fecal occult blood test, or at least one barium enema, sigmoidoscopy, or colonoscopy during the reporting period
Use of ACE inhibitor in CHF	Continuously enrolled members aged 18 years and older who had either two outpatient or one inpatient encounter with a diagnosis of CHF during the reporting period with exception of members without pharmacy benefits or members with at least one diagnosis of diastolic heart failure during the reporting year	Members in the denominator who filled at least one prescription for an ACE inhibitor or an angiotensin receptor blocker, or nitrates and hydralazine during the reporting period
Use of long-term asthma control drugs	Continuously enrolled members aged 5–64 years with moderate to severe persistent asthma during the reporting period, as evidenced by at least one emergency or inpatient encounter with a principal diagnosis of asthma, or at least six outpatient encounters with a principal diagnosis of asthma	Members in the denominator who received at least one prescription for a long-term asthma control drug during the reporting period
Diabetic retinal exam	Continuously enrolled members aged 18 years or older who were identified as having diabetes	Members in the denominator who have had an ophthalmoscopic examination performed by an eye care professional during the reporting period

*Continued*

Table 1: *Continued*

<i>Indicator</i>	<i>Denominator</i>	<i>Numerator</i>
Hemoglobin A1c for diabetics	Continuously enrolled members aged 18 years or older who were identified as having diabetes	Persons in the denominator who received at least two Hb A1c tests during the reporting period
Antihypertensive drug compliance	Continuously enrolled members aged 18 and older with hypertension and at least two prescriptions for an antihypertensive medication during the reporting period with exception of members without pharmacy benefits and those who had hypertension associated with pregnancy, childbirth, or the puerperium	Members in the denominator who received antihypertensive prescription coverage for at least 80% of eligible days during the reporting period
Lipid-lowering drug compliance	Continuously enrolled members aged 18 years and older with hyperlipidemia and at least two prescriptions for lipid-lowering medication during the reporting period with exception of members without pharmacy benefits	Members in the denominator who received lipid-lowering drug prescription coverage for at least 80% of eligible days during the reporting period
Childhood immunizations VZV	Continuously enrolled children whose second birthday fell during the reporting period, excluding members with contraindications for VZV	Members in the denominator who had VZV vaccination between the first and second birthdays or with a history of varicella disease
Childhood immunizations MMR	Continuously enrolled children whose second birthday fell during the reporting period, excluding members with contraindications for MMR	Members in the denominator who had MMR vaccination between the first and second birthdays or disease diagnosis for measles, mumps, and rubella

ACE, angiotensin-converting enzyme; CHF, congestive heart failure; VZV, varicella-zoster virus; MMR, measles, mumps, and rubella; PQSR, Physician Quality and Service Recognition.

Table 2: PQSR Program Highlights

Year	Budget (Million \$USD)	Award Amount as Percent of Base Professional Fees (%)	Individual Awards (Range, \$USD)	Number of Eligible Physicians	Number of Participating Physicians	Improvement Bonus (Range, \$USD)	Average Total Payment (\$USD)	Median Total Payment (\$USD)
1998	2.07	1.0-5.0	250-10,000	1,698	855	-	2,428	1,890
1999	2.64	1.0-5.0	250-10,000	2,164	1,198	-	2,216	1,508
2000	3.84	1.0-5.0	250-12,500	2,411	1,466	-	2,615	1,838
2001	4.35	1.0-5.0	250-12,500	2,455	1,548	0-3,000	2,800	2,092
2002	9.12	1.0-7.5	250-16,000	2,492	1,758	0-7,500	5,172	4,280
2003	9.70	1.0-7.5	250-16,000	2,613	2,030	0-8,000	4,785	3,744

PQSR, Physician Quality and Service Recognition.



close contact with physicians and local professional associations throughout the initial program implementation, as well as via satisfaction surveys, and an annual review of the program. Recent survey results indicated that more than half of participating physicians were “satisfied” or “very satisfied” with the concept of pay for performance. Also, Hawaii physicians and specialty associations participated in annual review of the quality indicators and provided feedback in the selection of indicators for the PQSR program. Transparency of the methods used to determine performance rates and translate the rates into a financial incentive amount was maintained via a public website.

### Data

Administrative claims data provided by HMSA were analyzed in this study. Data were collected annually for 6 years starting with the first year of program implementation (1998–2003). The overall enrolled PPO membership during the study period consisted of approximately one-half of a million members. The claims database contained membership information, pharmacy claims, and line-item inpatient, emergency department, and outpatient service claims. Specifically, the following data elements were available: member demographics, such as age, gender, and enrollment history; inpatient uniform billing (UB-92) data including principal and secondary *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9) diagnosis and procedure codes (ICD-9 2001), and admission and discharge dates; outpatient data including Health Care Financing Administration (HCFA 1500) information, principal and secondary ICD-9 diagnosis codes, *Current Procedural Terminology, Fourth Edition* (CPT-4) codes (AMA 1999), and date of service; and pharmacy data including generic and brand name, therapeutic class or generic product identifier, National Drug Code (NDC), and prescription fill date and days supply.

Poor data quality remains a typical concern of measuring clinical performance using administrative claims data (Greenberg 2001; Smith and Scanlon 2001). While the primary purpose of the claims data used in this study was provider reimbursement, several characteristics of this health plan’s database address some of the concerns including the availability of complete enrollment data for eligible members (primary subscribers as well as dependents) and comprehensive coverage of all services measured in the PQSR program quality indicators under the PPO benefit. In addition, given that physician services are reimbursed on a transaction basis in a PPO model, the utility of administrative claims data for analysis was likely increased, as compared with,

say, a capitated model. These characteristics applied to all claims data submitted by physicians to HMSA, hence, there was no indication that data collection quality varied between participating and nonparticipating providers.

### *Statistical Analysis*

The member was the unit of analysis. The study outcome—receipt of recommended care—was measured as a binary variable coded “1” if the member received recommended care for a specific indicator and “0” if the member did not receive recommended care. Hence, each member could have multiple records for each year as well as across years depending on the number of applicable indicators.

Members were stratified into three mutually exclusive groups based on physician visits during each year: those who visited only program-participating physicians, those who visited only nonparticipating physicians, and those who visited both participating and nonparticipating physicians in the same year. As only physicians of applicable specialties were evaluated for each quality indicator, visits to participating, visits to nonparticipating, or visits to both types of physicians were flagged only if the member saw a corresponding physician of applicable specialty. For example, visits to an obstetrician were considered whereas visits to a surgeon were not considered when determining applicable physician visits for the breast cancer screening indicator, even if those visits occurred during the measurement period. We chose this stratification to make one clean comparison between members who visited only participating physicians and members who visited only nonparticipating physicians, as well as between members who visited both types of physicians and members who visited only nonparticipating physicians, acknowledging that the latter relationship was likely to be weaker.

Time was included in the model using a discrete variable ranging from 0 to 5 for program years 1998–2003, respectively. A quadratic term was also included to account for the nonlinear trend in the data over time. The regressor of interest was the group by program year interaction terms, where the reference group was comprised of members with visits to only nonparticipating physicians. Other independent variables included member age and sex, and total number of physician office visits in a given program year. Total number of physician office visits and group assignment were treated as time-varying factors.

A generalized estimating equation (GEE) approach was used to account for the intraclass correlation between repeated observations for the same

individual (Liang and Zeger 1986). GEE is a nonparametric method that adjusts the variance using a working matrix of correlation coefficients to account for the correlation between observations and was employed using a logit link function and a binomial distribution assuming an autoregressive or independent correlation structure.

In an effort to address the potential bias of self-selection into the program at the physician level, the previous year's rates of performance between physicians who chose not to join and physicians who chose to join the program for the first time in a given year were compared for each indicator using the Satterthwaite *t*-test. Performance rates of the physicians who chose to join the program in the first year and the rates of the physicians who did not join the program in the first year were also compared.

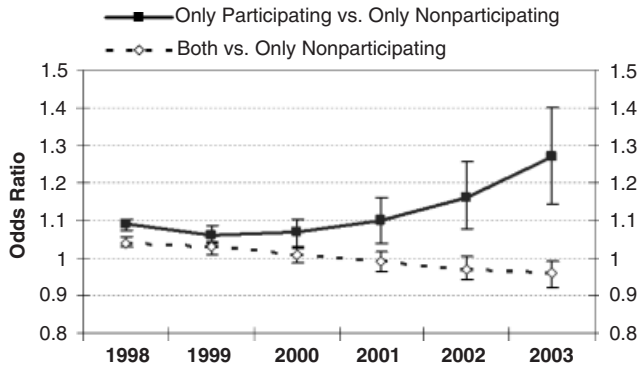
The *p*-values ( $<.05$ ) were considered statistically significant. SAS<sup>®</sup> Proprietary Software, Release 8.2 (SAS Institute Inc., Cary, NC) and STATA (Statacorp 2003) were used for all statistical analyses.

## RESULTS

The number of members eligible for at least one of the 11 indicators in the first program year was 222,213, which remained relatively stable over the study period. The proportion of eligible members who visited only program-participating physicians increased by 7 percent during the 6-year study period whereas the proportion of members who visited nonparticipating physicians dropped by 13 percentage points. The number of members visiting both program-participating and nonparticipating physicians increased by 6 percent over the study period.

Across all program years, members who visited only program-participating physicians had significantly higher odds (odds ratio [OR] = 1.06–1.27; 95 percent confidence interval [CI]: 1.03–1.08, 1.09–1.40) of receiving recommended care as measured by the 11 indicators compared with members who visited only nonparticipating physicians (Figure 1). For the members who visited both program-participating and nonparticipating providers, the odds of receiving recommended care decreased from 1.04 to 0.96 between 1998 and 2003, as compared with members who visited only nonparticipating providers. In general, members who were older or male were less likely to receive recommended care whereas members with more physician office visits were more likely to receive recommended care (Table 3).

Figure 1: Odds of Receiving Recommended Care over Time.



To account for potential selection bias, physician performance rates were compared between physicians who joined the program and those who did not join the program in the first year. In the following years, the previous year’s performance rates were compared for physicians who joined the program for the first time versus those who did not join the program in a given

Table 3: GEE Regression Coefficients Representing Change in Likelihood of Receiving Recommended Care as per 11 Quality Indicators\*

Variable	Coefficient <sup>†</sup>	Odds Ratio	95% Confidence Interval
Age	-0.015	0.99	-0.0152, -0.0149
Female	0.21	1.23	0.219, 0.209,
Only participating physicians	0.086	1.09	0.100, 0.072
Both participating and nonparticipating physicians	0.042	1.04	0.057, 0.029
Only nonparticipating physicians	Reference		
Total number of visits	0.284	1.33	0.281, 0.286
Year	0.059	1.06	0.051, 0.068
Quadratic year	-0.011	0.99	-0.012, -0.009
Year <sup>†</sup> only participating physicians	-0.039	0.96	-0.029, -0.049
Year <sup>†</sup> both participating and nonparticipating physicians	-0.017	0.98	-0.012, -0.022
Quadratic year <sup>†</sup> only participating physicians	0.014	1.01	0.016, 0.012

\*A total of 3,257,645 observations were included in the model. Clustering occurred at the patient level.

<sup>†</sup>All coefficients were significant at  $p < .001$ .

GEE, generalized estimating equations.

year. In the first program year, unadjusted comparisons of physician performance rates by indicator yielded similar rates for nine of the 11 indicators. Physicians who joined the program in the first year had significantly higher performance rates for use of ACE inhibitor in CHF (48 versus 42 percent,  $p = .003$ ) and breast cancer screening (57 versus 53 percent,  $p = .009$ ) compared with those who did not join the program.

In the subsequent years, physicians who joined the program for the first time in a given year had higher previous year's performance rates compared with physicians who did not join the program for the following indicators: colorectal cancer screening (27 versus 21 percent,  $p = .03$ ), cervical cancer screening (6 versus 58 percent,  $p = .03$ ), and breast cancer screening (59 versus 51 percent,  $p = .003$ ) for physicians who joined for the first time in 1999; cervical cancer screening (59 versus 53 percent,  $p = .04$ ) and MMR vaccine (88 versus 78 percent,  $p = .04$ ) for 2000; HbA1c testing among members with diabetes for 2001 (68 versus 56 percent,  $p = .02$ ); colorectal cancer screening (25 versus 20 percent,  $p = .03$ ), retinal exam for members with diabetes (48 versus 41 percent,  $p = .04$ ), and breast cancer screening (6 versus 53 percent,  $p = .001$ ) for 2002; and use of long-term control drugs for asthma (86 versus 78 percent,  $p = .04$ ) and breast cancer screening (62 versus 54 percent,  $p = .04$ ) for 2003. In 1999, performance rates for the retinal exam for members with diabetes indicator were significantly lower for physicians who joined the program in the following year compared with those who did not join (30 versus 38 percent,  $p = .01$ ).

## DISCUSSION

Recent studies found considerable evidence of serious deficiencies in the quality of care delivered to the U.S. population (Jencks et al. 2000; McGlynn et al. 2003; NCQA 2003), indicating an urgent need for an innovative solution. This patient-centered evaluation of a quality-based physician incentive program implemented in a PPO setting revealed a positive association between the quality of care delivered to plan members and whether they received care from a program-participating physician. Moreover, the association grew stronger in later years when members who visited only program-participating physicians were compared with members who visited only nonparticipating physicians. Altogether, these findings suggest that a quality-based financial incentive program aimed to improve the quality of care delivered to health plan members through differential quality-based reimbursement payments

to the participating physicians can significantly improve the quality of care provided to the member population, as measured by these 11 indicators.

We observed a negative trend in the association between the quality of care and the group of patients who saw both PQSR-participating and non-participating physicians. While the association was nonsignificant in 2000–2002, the meaning of this finding is unclear. It is possible that the trend is a by-product of increasing physician participation in the program over time. Physician participation increased from about 50 percent in the first year to 78 percent in year 2003. As more physicians became PQSR-participants, it was increasingly more likely that patients who saw both participating and non-participating providers were the patients encountering the health care system more frequently than those in the “only” groups or those in the “both” group in the prior year. It is possible that these patients who saw both types of providers (participating and nonparticipating) were either receiving less coordinated care or had a greater disease burden compared with the patients in the “only” groups or compared with those in the “both” group in the previous years. Future studies will assess performance of the “both” group more closely over time in light of this hypothesis.

While physician performance on the quality indicators is the major component of the incentive program, other components of the program include board certification, patient satisfaction, business operations (e.g., use of electronic records), and measures of efficiency. This analysis evaluated whether or not there has been an effect associated with the overall program, but does not evaluate specific mechanism of influence. Further research is necessary to isolate the relative impact of the quality indicators versus other components of the program as well as the effect of financial incentives versus educational or awareness aspects of the program (Chassin 2005). For instance, several of the indicators used in the program overlap with the HEDIS<sup>®</sup> measures, which record an overall positive trend toward better performance and awareness of quality issues (NCQA 2003). In general, however, our approach accounts for any systemic events during the 6 years after program implementation that equally affected both the intervention and comparison groups (such as disease management or reminder programs implemented across the health plan population).

Data were not available for years before program implementation, precluding a comparison of pre- and postimplementation trends. Hence, we could not evaluate whether the trends observed postimplementation existed before the program was implemented. In addition, while several characteristics of this claims database suggest that the data were generally complete,

certain potential biases inherent in claims data should be mentioned. These include coding variation between providers, time lag between receipt of service and claim processing, and missing data, especially in fields not relevant to reimbursement or potentially sensitive information due to confidentiality concerns (Smith and Scanlon 2001). Furthermore, vaccination claims might be underreported because of free vaccine serum supplies from public sources (Levin-Scherz, DeVita, and Timbie 2005). Generally, however, only those factors that differentially affected the intervention and comparison groups could bias the results.

The potential effect of selection bias was considered as physicians who typically follow recommended guidelines might have chosen to enroll in the program each consecutive year thereby mitigating a detectable overall change in the rates of recommended care delivery (Chassin 2005). Although the real-world naturalistic set-up of the program made it difficult to control for this potential selection bias, a separate analysis was performed to determine whether better performing physicians were self-selecting into the program. The results revealed that for a majority of the indicators, physicians who joined the program in a given year did not perform significantly better in the previous year than the physicians who decided not to join the program. Less than a quarter of eligible physicians chose not to participate in the incentive program in 2003. Implementing mandatory program participation for all physicians contracting with the health plan would likely address the issue of selection bias.

It is important to place these findings in light of other evaluations of pay-for-performance programs. Unfortunately, most interventions of pay-for-performance programs have been conducted in the HMO setting (Levin-Scherz, DeVita, and Timbie 2005; Rosenthal et al. 2005), which is characterized by several important distinctions from the PPO setting. The current study is unique in that it was implemented in a PPO plan, where members were not assigned a “gate-keeper” primary care provider. This important characteristic of the current study may have strong influence on the results, as the base compensation arrangements in PPO plans may in fact reinforce the pay-for-performance incentives that encourage appropriate utilization of certain services (e.g., follow-up, laboratory test, etc.) (Dudley 2005). Furthermore, the PPO plan structure complicated physician attribution of patient care, dictating a more “generous” attribution. Because patients had free access to any PPO network physician without referral, any physician of applicable specialties who came into contact with the patient was credited for the care the patient received, whether or not the physician actually provided the care. Further

analysis of the extent of such “cross-subsidizing” within this attribution arrangement and its impact on quality of care is in order.

Rosenthal et al. (2005) observed similar improvement in clinical quality scores on select three indicators for both a physician network that offered a financial reward and a contemporaneous physician network that did not receive an incentive. Within the pay-for-performance network, physician groups with higher performance at baseline received a disproportional larger percentage of the overall incentive budget (Rosenthal et al. 2005). While we similarly observed positive trends both for patients with visits to nonparticipating and for patients with visits to both participating and nonparticipating providers, our analysis controlled for any baseline temporal trends by including a local control group (nonparticipating providers). Levin-Scherz, DeVita, and Timbie (2005) observed mixed results in an HMO-based incentive program focused on diabetes and asthma measures: improvement was observed for some of the indicators compared with regional and national benchmarks. As in the Rosenthal study, the incentive program employed network-wide performance targets and measured outcomes at the physician-group level. The PQSR program evaluated in this study rewarded superior performance relative to provider peers within the program, but also included an improvement bonus to reward improved performance compared with previous year. This key difference in the study design may also help explain the difference in results generated by this study. Furthermore, our study employed a local control group with similar characteristics of health plan and patient populations in both control and treatment group, longer follow-up time, and a more extensive/different set of quality indicators, which are all important factors affecting outcomes (Dudley 2005).

The rate of indicator compliance varied considerably across indicators at baseline. For example, 53 percent of eligible women were receiving a screening mammogram in 1998, whereas 69 percent of members with diabetes were receiving at least one HbA1c test in the same year. As a result, one might expect the aggregate effect of these indicators to be somewhat diluted, given different capacities for improvement. Despite these indicator-specific limitations, a consistent, positive association between visiting program-participating physicians and receiving recommended care was demonstrated for this PPO plan member population at the aggregate level during each of the 6 program years, thus supporting the study findings and conclusions.

While this evaluation provides evidence of the utility of the incentive program in positively affecting the quality of health care delivery over time, further evaluation of the individual indicators should be conducted to



determine whether some indicators are more effective than others. In addition, although this program was designed and implemented under real-world conditions, further analysis is needed to test the generalizability of this program to other populations. While this plan's position as the largest provider of health care coverage in the state of Hawaii improved its ability to influence physician behavior, it is unclear whether the same results could be expected by a health plan that does not contribute as large a share of physician compensation. Also, a study examining the amount of financial incentive that is most effective for achieving targeted levels of recommended patient care remains an important area of future research.

Furthermore, issues pertinent to any quality-based performance program must be considered and warrant careful interpretation of the findings. Performance-based metrics generally do not fully consider the member's severity of disease and individual preferences for tests and treatment when evaluating actions taken by the provider (Walter et al. 2004). Program performance rates for each physician may be biased with respect to the physician's patient case mix. The program uses relative indicator-specific physician ranking in bonus amount calculations, rather than implementing strict system-wide target rates. The relative ranking approach accounts for indicator-specific patient compliance and severity of disease differences, but does not account for any differences specific to a physician's patient case mix.

Patient compliance is an important factor in the measurement of physician quality based on receipt of care by the patient. The low compliance rates of certain indicators such as colorectal and cervical cancer screening across years were probably due in large part to patient preference that outweighed physician recommendations. Quality indicators that require a visit to a specialist (e.g., retinal exam for members with diabetes) or additional payment for prescription medication (e.g., continuity of supply of antihypertensive drugs) may also exhibit a stronger association with patient preferences due to the increased costs involved in receiving care. Such a phenomenon would likely result in a type II error, which would in turn dilute the effect of the physician incentive program. Further analysis of both high- and low-preference indicators in a context that examines them individually and directly as well as predictors of noncompliance is therefore warranted.

In conclusion, the results of this study suggest that the improvement in the delivery of recommended care, as measured by the 11 evidence-based performance indicators, achieved during the study period may be a result of the physician incentive program. Notwithstanding the limitations of evaluating the effects of a quality-based incentive program in a real-world setting, it is

important to recognize the wide dissemination and interest for this new breed of physician reimbursement models by some of the leading private and public payers in the United States. The concept of reimbursing providers based—at least in part—on the quality of care is not only a novel approach that is gaining popularity within the health care sector, but an innovation that may have the potential to improve the quality of care. Further study of the effectiveness of such models is necessary to fully take advantage of the considerable promise of this innovation to address the objective of improving the quality of health care in the United States.

## ACKNOWLEDGMENTS

This study was supported by the Hawaii Medical Service Association, Honolulu, Hawaii.

*Disclosures:* Drs. Chung and Taira are employees of Hawaii Medical Service Association, the health plan whose physician incentive program was analyzed in this study. Ms. Gilmore, Kang, Ryskina, Dr. Zhao, and Dr. Legorreta report no conflicts of interest.

*Disclaimers:* Hawaii Medical Service Association reviewed the manuscript before submission for publication and contributed considerable information about the program and administrative data for analysis, but had no influence over study design, analysis, or manuscript write-up.

## REFERENCES

- Adams, K., and J. M. Corrigan. 2003. *Priority Areas for National Action: Transforming Health Care Quality*. Committee on Identifying Priority Areas for Quality Improvement. Institute of Medicine of the National Academies Washington, DC: National Academies Press [accessed on March 21, 2007]. Available at [www.nap.edu](http://www.nap.edu)
- American Medical Association (AMA). 1999. *Current Procedure Terminology*, 4th edition. Chicago: American Medical Association.
- Amundson, G., L. I. Solberg, M. Reed, E. M. Martini, and R. Carlson. 2003. "Paying for Quality Improvement: Compliance with Tobacco Cessation Guidelines." *Joint Commission Journal on Quality and Safety* 29 (2): 59–65.
- Carlisle, D. M., A. L. Siu, E. B. Keeler, E. A. McGlynn, K. L. Kahn, L. V. Rubenstein, and R. H. Brook. 1992. "HMO vs. Fee-for-Service Care of Older Persons with Acute Myocardial Infarction." *American Journal of Public Health* 82 (12): 1626–30.

- Casalino, L., R. R. Gillies, S. M. Shortell, J. A. Schmittiel, T. Bodenheimer, J. C. Robinson, T. Rundall, N. Oswald, H. Schaufli, and M. C. Wang. 2003. "External Incentives, Information Technology, and Organized Processes to Improve Health Care Quality for Patients with Chronic Diseases." *Journal of the American Medical Association* 289 (4): 434-41.
- Center for Medicare and Medicaid Services (CMS). 2003. "Quality Initiatives. Doctors Office Quality Project" [accessed on September 17, 2003]. Available at <http://www.cms.hhs.gov/quality>
- Chassin, M.R. 2005. "Does Paying for Performance Improve Quality of Health Care?" *Medical Care Research and Review* 63 (1): 122S-5S.
- Chung, R. S., H. O. Chernicoff, K. A. Nakao, and R. C. Nickel. 2004. "A Quality-Driven Physician Compensation Model: Four-Year Follow-up Study." *Journal for Healthcare Quality* 25 (6): 31-7.
- Corrigan, J. M., J. Eden, and B. M. Smith. 2002. *Leadership by Example: Coordinating Government Roles in Improving Health Care Quality*. Committee on Enhancing Federal Healthcare Quality Programs. Institute of Medicine of the National Academies Washington, DC: National Academies Press [accessed on March 21, 2007]. Available at [www.nap.edu](http://www.nap.edu)
- Dudley, R.A. 2005. "Pay-for-Performance Research: How to Learn What Clinicians and Policy Makers Need to Know." *Journal of the American Medical Association* 294 (14): 1821-3.
- Fairbrother, G., K. L. Hanson, S. O. Friedman, and G. C. Butts. 1999. "The Impact of Physician Bonuses, Enhanced Fees, and Feedback on Childhood Immunization Coverage Rates." *American Journal of Public Health* 89: 171-5.
- Forsberg, E., R. Axelsson, and B. Arnetz. 2001. "Financial Incentives in Health Care. The Impact of Performance-Based Reimbursement." *Health Policy* 58 (3): 243-62.
- Greenberg, L. 2001. "Overview: PPO Performance Measurement: Agenda for the Future." *Medical Care Research and Review* 58 (1): 8-15.
- Hellinger, F.J. 1998. "The Effect of Managed Care on Quality: A Review of Recent Evidence." *Archives of Internal Medicine* 158 (8): 833-41.
- ICD-9. *International Classification of Diseases, Ninth Revision, Clinical Modification*. 2001. Salt Lake City, UT: Medicode Publications.
- Institute of Medicine (IOM). 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Committee on Quality of Health Care in America: Institute of Medicine. Institute of Medicine of the National Academies Washington, DC: National Academies Press [accessed on March 21, 2007]. Available at [www.nap.edu](http://www.nap.edu)
- Jencks, S. F., T. Cuerdon, D. R. Burwen, B. Fleming, P. M. Houck, A. E. Kussmaul, D. S. Nilasena, D. L. Ordin, and D. R. Arday. 2000. "Quality of Medical Care Delivered to Medicare Beneficiaries: A Profile at State and National Levels." *Journal of the American Medical Association* 284: 1670-6.
- Kohn, L. T., J. M. Corrigan, and M. S. Donaldson. 2000. *To Err Is Human: Building a Safer Health System*. Committee on Quality of Health Care in America: Institute

- of Medicine of the National Academies Washington, DC: National Academies Press [accessed on March 21, 2007]. Available at [www.nap.edu](http://www.nap.edu)
- Levin-Scherz, J., N. DeVita, and J. Timbie. 2005. "Impact of Pay-for-Performance Contracts and Network Registry on Diabetes and Asthma HEDIS Measures in an Integrated Delivery Network." *Medical Care Research and Review* 63 (1): 14S-28S.
- Liang, K. Y., and S. L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73: 13-22.
- McGlynn, E. A., S. M. Asch, J. Adams, J. Keesey, J. Hicks, A. Decristofaro, and E. A. Kerr. 2003. "The Quality of Health Care Delivered to Adults in the United States." *New England Journal of Medicine* 348: 2635-45.
- Merrill, R. M., M. L. Brown, A. L. Potosky, G. Riley, S. H. Taplin, W. Barlow, and B. H. Fireman. 1999. "Survival and Treatment for Colorectal Cancer: Medicare Patients in Two Group/Staff Health Maintenance Organizations and the Fee-for-Service Setting." *Medical Care Research Review* 56 (2): 177-96.
- Morrow, R. W., A. D. Gooding, and C. Clark. 1995. "Improving Physician's Preventive Health Care Behavior through Peer Review and Financial Incentives." *Archives of Family Medicine* 4 (2): 165-9.
- National Committee for Quality Assurance (NCQA). 2003. "GE, Ford, UPS, P&G, Verizon, Others Back New Pay-for-Quality Initiative for Physicians." *NCQA News*.
- . 1999-2001. *HEDIS<sup>®</sup> Technical Specifications*. Washington, DC: NCQA.
- Rosenthal, M. B., R. G. Frank, Z. Li, and A. M. Epstein. 2005. "Early Experience with Pay-for-Performance: From Concept to Practice." 294 (14): 1788-93.
- Roski, J., R. Jeddelloh, L. An, H. Lando, P. Hannan, C. Hall, and S. H. Zhu. 2003. "The Impact of Financial Incentives and a Patient Registry on Preventive Care Quality: Increasing Provider Adherence to Evidence-Based Smoking Cessation Practice Guidelines." *Preventive Medicine* 36 (3): 291-9.
- Smith, D. G., and D. P. Scanlon. 2001. "Covered Lives in PPOs." *Medical Care Research and Review* 58 (1): 16-33.
- Sommers, A. R., and D. R. Wholey. 2003. "The Effect of HMO Competition on Gatekeeping, Usual Source of Care, and Evaluation of Physician Thoroughness." *American Journal of Managed Care* 9 (9): 618-27.
- StataCorp. 2003. *Stata Statistical Software: Release 8*. College Station, TX: StataCorp LP.
- Walter, L., N. Davidowitz, P. Heineken, and K. Covinsky. 2004. "Pitfalls of Converting Practice Guidelines into Quality Measures." *Journal of the American Medical Association* 291 (20): 2466-70.
- Webber, A. 2005. "Pay for Performance: National and Local Accountability for Health Care." *Managed Care* 14 (12, suppl): 3-5.