

## Clinical Quality Indicators and Provider Financial Incentives: Does Money Matter?

Kathleen A. Fairman, MA, and Frederic R. Curtiss, PhD, RPh, CEBS

*“Simply put, if reimbursement can drive utilization and utilization can drive outcome, reimbursement can drive outcome.”<sup>1</sup>*

In the decades-old debate over how best to finance and deliver health care in the United States, a nearly ubiquitous complaint about most systems of physician and hospital reimbursement is that payments are made based on the *services* delivered regardless of the *quality of care* delivered, providing no incentive—some say a disincentive—for quality improvement.<sup>2,3</sup> Advocates of “pay-for-performance” (P4P) systems of health care reimbursement argue that the concept of paying more to those who produce better outcomes, a “bedrock principle” in efforts to “reduce error and reinforce best practices” in other industries, should become “a top national priority” in “the campaign to rally our underperforming health care system.”<sup>3</sup>

In proposals to improve health care systems, high-level enthusiasm is not necessarily an indicator of high-quality evidence, and P4P is no exception. Editorialists George Diamond and Sanjay Kaul, both cardiologists and keen observers of quality of evidence in health care decision making, wrote in 2009 that rapid proliferation of P4P systems “is occurring despite a paucity of empirical evidence that [they] actually deliver on their promise to improve the quality and reduce the cost of health care. There are essentially no randomized controlled trials (RCTs) demonstrating the effectiveness of [P4P] programs and very few reports in the literature that analyze the existing programs.”<sup>1</sup> The point made by Diamond and Kaul is well-taken. As we have observed previously, the health care research literature is replete with examples of schemes that were widely (sometimes wildly) supported based on weak observational evidence but refuted and ultimately abandoned after being tested with more rigorous research designs.<sup>4</sup> When considering implementation of a P4P program, managed care organization (MCO) decision makers should be mindful *both* of quantitative research findings about the degree to which P4P and similar interventions improve quality *and* of the limitations of the evidence base at the present time. Key issues are (a) whether improvements associated with P4P would have taken place anyway, that is, without the financial incentives and (b) what factors “drive” changes in provider behavior.

### Effects of Quality Improvement Interventions Alone and with Financial Incentives

In this issue of *JMCP*, Brackbill et al. report the results of a

quality improvement project undertaken to increase the percentage of patients receiving discharge orders for chronic aspirin therapy following a hospitalization for acute myocardial infarction (AMI) or coronary artery bypass graft (CABG).<sup>5</sup> Using a pre-intervention versus post-intervention study design, Brackbill et al. found that an intervention consisting of provider education coupled with the placement of a colorful “prescription” for aspirin in the inpatient chart of patients clinically eligible for aspirin therapy was associated with a change in the aspirin discharge order rate from 94.9% pre-intervention to 98.9% post-intervention ( $P=0.012$ ). When analyzed by subgroup, the relationship between the intervention and aspirin discharge order rate was statistically significant for patients hospitalized for CABG (change from 91.5% to 100.0%,  $P=0.016$ ) but not for AMI (change from 96.6% to 98.5%,  $P=0.263$ ).

The efforts of Brackbill et al. in using a novel approach to improve an important quality metric are commendable. Nonetheless, their results illustrate the challenges experienced by providers and payers that try to “move the needle” of compliance with treatment guidelines, especially when baseline compliance is high. Brackbill et al. report that quality assurance audits conducted shortly after the start of the project revealed substantial implementation problems; the aspirin “prescription” had been placed in only 25% of the charts of clinically eligible patients. Education of providers that was initially conducted pre-intervention had to be repeated twice in the project’s post-implementation phase.

It is noteworthy that the measure for which no significant difference was found following the intervention described by Brackbill et al. was the well-known “AMI-2,” the proportion of AMI patients with aspirin prescribed at discharge. AMI-2 has been used by the Centers for Medicare & Medicaid Services (CMS) in multiple hospital quality-of-care demonstrations and reimbursement incentive programs.<sup>6</sup> These include the Reporting Hospital Quality Data for Annual Payment Update program that was mandated by the Medicare Modernization Act of 2003 to allow CMS to “to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates”<sup>7</sup> and the CMS/Premier Hospital Quality Incentive demonstration (HQID), instituted in October 2003 as part of CMS efforts to “improve the care provided by the nation’s hospitals and to provide quality information to consumers and others.”<sup>8,9</sup> For hospitals and physicians, efforts

such as that undertaken by Brackbill et al. increasingly represent not only an attempt to improve patient care, but also a factor in the financial “bottom line.”<sup>9,10</sup>

### Evidence in Support of Using Financial Incentives to “Drive” Behavioral Change

Observational evidence generally suggests a relationship between financial incentives and documented compliance with quality outcome measures. For example, in the first 4 years of the HQID, which provides Medicare payment bonuses to hospitals that perform well on quality measures for patients with AMI, CABG, heart failure, pneumonia, and hip or knee replacement, bonuses totaling more than \$36.5 million were awarded to top performers.<sup>9</sup> During that time, care for all 5 clinical conditions improved, dramatically for a few conditions, among participating hospitals. For example, the CMS composite quality score for care provided to patients with heart failure improved from 64.5% to 92.2%; scores for pneumonia patients improved from 69.3% to 92.6%, and scores for AMI patients from 87.5% to 96.3%.<sup>9</sup> In an August 2009 press release summarizing the results of multiple similar financial incentive programs, CMS reported that “Medicare demonstrations... provide strong evidence that offering financial incentives for improving or delivering high quality care increases quality and can reduce the growth in Medicare expenditures” by finding that providers who choose participation in a quality incentive demonstration project improve their quality scores over time.<sup>11</sup>

Observational results for the private sector are similar. In a 10-year study of 2 quality indicators, screening for diabetic retinopathy and cervical cancer, Lester et al. (2010) found an association between financial incentives paid by an MCO to medical facilities and compliance with recommended screening protocols.<sup>12</sup> During 5 years (1999-2003) when facilities were paid bonuses for higher rates of diabetic retinopathy screening, the screening rate increased from 84.9% to 88.1%. Following a subsequent 4-year period in which the incentive was removed, the rate fell to 80.5%. Results were similar, but notably modest, for cervical cancer screening. During an incentive period from 1999-2000, the screening rate increased slightly from 77.4% to 78.0%. After a 5-year period without financial incentives, the rate declined to 74.3%.<sup>12</sup>

In a multispecialty group practice, Chung et al. (2010a) used a pre-implementation versus post-implementation design to study the addition of financial incentives in 2007 to a system of monitoring and quarterly reporting of quality indicators that had been in place for the previous 4 years.<sup>13</sup> Physicians could earn up to \$5,000 annually for compliance with 9 indicators of health care quality, including 3 clinical outcome measures for patients with diabetes (control of hemoglobin A1c, blood pressure, and low-density lipoprotein cholesterol [LDL-C]); use of asthma controller medication by patients in the system’s

asthma registry; screening for cervical cancer, *Chlamydia*, and colon cancer; and documentation of tobacco use history and of height and weight. In the incentive period beginning in 2007 (n=167 physicians) compared with 2006 (n=169), quality scores improved modestly for 8 of the 9 measures, but declined slightly (from 63% to 60%) for LDL-C control among patients with diabetes, reportedly because of a systematic problem with new laboratory equipment.<sup>13</sup> The largest percentage point increase was for the colon cancer screening rate, which increased from 40% to 47%; all other increases were less than 5 percentage points.<sup>13</sup>

Analyses of this type, although providing valuable descriptive information, beg the question of whether quality would have improved without the financial incentives—that is, how much does quality improvement represent “secular trend,” (changes over time unrelated to the intervention) or the effects of spurious factors, such as education provided simultaneously with financial incentives?

### Differentiating Secular Trend from Incremental Change Associated with Financial Incentives

Despite the common belief that financial incentives are necessary to drive health care providers to do the right thing to improve quality of care, it is important to ask whether P4P adds incrementally to trends attributable to other factors such as the promulgation of treatment guidelines or widespread use of auditing and feedback mechanisms; such factors potentially affect both clinical outcomes assessed by common quality measures and better documentation of those outcomes. In a longitudinal analysis of 42 primary care practices in 6 geographic regions of England, Campbell et al. (2007) addressed the key question of the incremental effect of adding provider financial incentives onto existing quality improvement initiatives.<sup>14</sup> Specifically, mean scores for 30 quality indicators with financial incentives were compared with mean scores for 17 quality indicators without financial incentives, first in 1998 and 2003 when the financial incentives were not yet effective and again in 2005 after the financial incentives had been in effect for 1 year. For 2 of the 3 clinical care categories, asthma and type 2 diabetes, the actual mean quality scores in 2005 exceeded the scores predicted by the trend in performance that was recorded in 1998 and 2003, suggesting that the incentives were associated with incremental improvement. However, there was no significant difference in the rate of improvement between quality scores with and without financial incentives.

### It’s Tough to Move a High Bar Higher: P4P Does Not Guarantee Meaningful Quality Improvement

Research employing comparison groups tells a somewhat different story about P4P than do simple pre-intervention versus post-intervention analyses, suggesting that the gains associated with (and often attributed to) P4P are small and inconsistent,

achieved by many providers without financial incentives, and especially modest among providers with high baseline compliance rates. For example, in an analysis of data originally collected for the Can Rapid Risk Stratification of Unstable Angina Patients Suppress Adverse Outcomes with Early Implementation of the American College of Cardiology/American Heart Association (ACC/AHA) Guidelines (CRUSADE), Glickman et al. (2007) compared 54 hospitals under P4P as part of the HQID with 446 non-HQID hospitals.<sup>15</sup> The subject of the comparison was the proportion of patients hospitalized for acute non-ST-segment elevation myocardial infarction who received the care specified by 6 quality indicators in 2003 and 2006. The 6 indicators included aspirin and beta-blocker at arrival and at discharge, angiotensin-converting enzyme (ACE) inhibitor or angiotensin II receptor blocker (ARB) at discharge for patients with left ventricular systolic dysfunction (LVSD), and counseling for smoking cessation for active or recent smokers. Baseline compliance in both groups exceeded 70% for all measures, 80% for 4 measures, and 90% for 2 measures. Percentage improvement over the 3 years for the composite of the 6 indicators did not significantly differ in the HQID hospitals (median 87.0% in 2003 versus 94.2% in 2006) compared with improvement in the non-HQID hospitals (88.0% to 93.6%,  $P=0.16$ ). For 4 of the 6 indicators— aspirin at arrival, beta-blocker at arrival, beta-blocker at discharge, and ACE inhibitor or ARB for patients with LVSD—rates of improvement for HQID and non-HQID hospitals did not significantly differ. Modest between-group differences in improvement rates from 2003 to 2006 were noted for 2 of the specific measures, aspirin at discharge (from 91.1% to 97.1% in HQID, from 92.2% to 95.9% in non-HQID,  $P=0.04$ ) and smoking cessation counseling (from 75.8% to 95.8% in HQID, from 74.0% to 88.8% in non-HQID,  $P=0.05$ ). Despite the significance of the between-group difference for the aspirin at discharge measure, the practical importance of a median proportion of approximately 97% versus 96% is questionable; the absolute 7% difference for smoking cessation is more clinically meaningful.

In a similar analysis, Lindenauer et al. (2007) used data reported as part of the Hospital Quality Alliance (HQA) project, a data collection effort in which 98% of hospitals nationwide report quarterly on a minimum of 10 quality indicators for 3 clinical conditions (heart failure, pneumonia, and AMI). Investigators compared 207 HQID hospitals with 406 non-HQID hospitals.<sup>16</sup> HQID and non-HQID hospitals were matched on size, teaching status, region, urban versus rural location, and nonprofit versus for-profit ownership; thus, the matching study hospitals represented a small subgroup of all HQA facilities that submitted sufficient data for all indicators ( $n=2,490$ ). For each of 3 medical conditions, the outcome measure was a “composite process score” calculated by dividing the number of patients who received “correct care” by the total number of treated patients. As in the Glickman et al. analysis,

baseline (fourth quarter of 2003) rates of compliance with the 5 quality indicators for patients with AMI were high in both the HQID and non-HQID groups, exceeding 85% for beta-blocker on arrival and at discharge, and exceeding 92% for aspirin on arrival and at discharge. Baseline compliance with the ACE inhibitor for LVSD indicator was 78.2% in HQID and 82.9% in non-HQID facilities.

Not surprisingly, Lindenauer et al. found a strong *inverse* relationship between baseline compliance and degree of improvement from baseline to follow-up (third quarter 2005); that is, compared with hospitals performing poorly at baseline, high-performing hospitals showed far less meaningful rates of improvement, even slightly declining in performance on some measures. In a summary composite measure that included all 3 study conditions, improvement for hospitals in the lowest baseline quintile was an absolute 16.1% (from 69.7% to 85.8%) for HQID and an absolute 13.2% (from 70.5% to 83.7%) for non-HQID ( $P=0.08$  for between-group comparison using paired t-test). In contrast, changes for hospitals in the highest baseline quintile were an absolute 1.9% (from 91.2% to 93.1%) in HQID and a decline of 3.0% (from 94.2% to 91.2%) for non-HQID ( $P<0.001$  for between-group comparison). Similar findings were noted for the AMI composite score; among HQID hospitals in the highest baseline quintile for AMI, performance was 97.9% at baseline and 96.8% at follow-up. For aspirin at discharge, a measure for which baseline compliance was 93.9% in HQID and 92.7% in non-HQID, improvements from baseline to follow-up were small (absolute 1%-2%) in both groups and did not significantly differ ( $P=0.48$ ). On the other hand, results for the HQID program overall were generally favorable, albeit modest. After statistical adjustment for “baseline performance and other hospital characteristics,” Lindenauer et al. reported that participation in the HQID was associated with composite measure improvements of 2.6% for AMI, 4.1% for heart failure, and 3.4% for pneumonia.<sup>16</sup>

A study of the effects of P4P payments made by a preferred provider organization to physicians for high-quality diabetes care was reported by Chen et al. (2010).<sup>17</sup> The investigators assessed the effects of financial incentives of 1.5%-7.5% of base professional fees paid from 1999 through 2006, capped to an annual maximum of \$10,000 to \$16,000 per physician but supplemented beginning in 2001 with an annual additional \$6,000 bonus paid to physicians who improved their performance compared with the previous year. Among patients with diabetes, those who received at least 2 A1c tests and 1 LDL-C test during the year were defined as having received high-quality care. For the sample overall, including patients of P4P participants and patients whose physicians declined P4P participation, rates of quality care increased from 42.3% of 19,573 patients in 1999 to 67.1% of 32,365 patients in 2006. During that time period, the percentage of patients seeing P4P-participating physicians increased as well, from 78.7% to 94.6%. In a random effects logit model controlling for demographic

and clinical characteristics (insulin dependence, receipt of care from an endocrinologist, comorbidity index, and number of primary care physicians seen during the year), patients of P4P participants were more likely to receive high-quality care (odds ratio [OR]=1.16, 95% confidence interval [CI]=1.11-1.22,  $P<0.001$ ) than were patients of physicians who declined P4P participation. Patients who saw a P4P-participating physician continuously from 2004 through 2006 had a lower all-cause hospitalization rate in 2006 than patients who did not (negative binomial modeling, incident rate ratio [IRR]=0.75, 95% CI=0.61-0.93,  $P<0.01$ ). However, all-cause hospitalization rates for patients of P4P and non-P4P physicians overall did not differ (IRR=1.00, 95% CI=0.95-1.05,  $P=0.27$ ).

### Randomized Trials of P4P: Limited Studies with Mixed Results

A PubMed search on the term “pay-for-performance” limited to RCTs yields only a handful of small studies. An et al. (2008) found that 24 clinics that were paid \$5,000 for 50 “quitline” referrals for smoking cessation referred a higher percentage of smokers than did 25 control group clinics (11.4% vs. 4.2%,  $P=0.001$ ).<sup>18</sup> Fairbrother et al. (2001) found that incentives provided to 57 “randomly selected inner-city physicians” produced changes in *documentation* of immunization for “50 randomly selected children,” but no significant change in the actual immunization rate.<sup>19</sup> Hillman et al. (1998) studied compliance with cancer screening guidelines (mammography, breast exam, colorectal screening, and Papanicolaou testing) for women aged 50 years or older in a Medicaid health maintenance organization (HMO), finding that 26 clinics randomized to receive written feedback and financial incentives from 1993 to 1995 did not significantly differ from 26 control clinics.<sup>20</sup> However, screening rates improved over time in both the intervention and control groups. Chung et al. (2010b) randomized providers who were receiving incentive payments of up to \$5,000 annually to be paid either a single annual lump sum or 4 quarterly payments, finding no significant between-group differences in quality scores or total incentive payments.<sup>21</sup>

The most important implication to come from these studies is not the information that they provide; it is the remarkably small amount of high-quality and relevant evidence that is available today to MCO decision makers who are assessing the potential merits of a financial incentive program for providers. Although a comprehensive literature review by Petersen et al. (2006) of studies published from 1980 through 2005 identified another handful of randomized studies of P4P, none assessed ambulatory care provided after the 1990s, and most were small-scale.<sup>2</sup> Petersen et al. deemed as “critical” a research agenda “including [RCTs] and observational studies with concurrent control groups,” consistent with the observations of Diamond and Kaul about the notable weaknesses in the current base of evidence about P4P.

### Limitations of Typical P4P Schemes: Can We Do Better?

To be successful, a system intended to promote improvement in the outcomes of care provided by physicians must be reasonably consistent with, or at least take into account, the way that physicians have been trained to practice medicine. In their editorial, Diamond and Kaul observe that the reporting of aggregate outcomes, a “key design feature” of most P4P systems, “flies in the face of everything we know about the actual practice of medicine. As any practicing physician will attest, the expected benefit associated with a particular therapy varies widely from patient to patient, and a fundamental part of the physician’s job is to determine – based on one’s clinical experience and one’s grasp of the current medical literature – which treatment is most appropriate for the patient at hand.”<sup>11</sup> The implication of Diamond and Kaul’s observation is that the most successful financial incentive schemes will be synergistic with—rather than fighting against—physicians’ training, rewarding those who do especially well at the important job of targeting treatment to each patient.

With this principle as backdrop, Diamond and Kaul make the provocative suggestion that in lieu of providing retrospective incentive payments based on aggregate outcomes, payers should turn to “evidence-based reimbursement,” in which the payment amount for a given procedure or medication would be based on the expected benefit for the patient, defined using the results of RCTs. For example, Diamond and Kaul cite the results of the Clinical Outcomes Utilizing Revascularization and Aggressive druG Evaluation (COURAGE) trial, which randomized patients with “objective evidence of myocardial ischemia and significant coronary artery disease” to medical therapy plus percutaneous coronary intervention (PCI,  $n=1,149$ ) or medical therapy alone ( $n=1,138$ ).<sup>22</sup> During a median follow-up of 4.6 years (range 2.5 to 7.0 years), cumulative rates of the primary event (all-cause death or nonfatal myocardial infarction) were 19.0% for the PCI group and 18.5% for the medical therapy group (hazard ratio [HR]=1.05, 95% CI=0.87-1.27,  $P=0.62$ ).<sup>22</sup> Results on secondary endpoint measures, including hospitalization for acute coronary syndrome or myocardial infarction, were similar. COURAGE investigators concluded that “in patients with stable coronary artery disease,” PCI did not confer significant improvement to patient outcomes compared with “optimal medical therapy” (defined as “intensive pharmacologic therapy and lifestyle intervention”) alone.

Based on the COURAGE results and physiologic evidence about the primary causes of atherosclerotic events, Diamond and Kaul suggest that PCI is “considered formally appropriate” only in patients who meet 3 criteria: (a) ischemic symptoms, (b) objective evidence of ischemia provided by stress testing, and (c) failure of an adequate trial of medication management. Observing that a large proportion of patients undergoing PCI do not meet these criteria, Diamond and Kaul suggest a

payment system in which the highest level of reimbursement for PCI would be made when the patient meets all 3 criteria; reimbursement would be reduced somewhat for PCI provided to patients meeting only 2 criteria; reduced further for patients meeting only 1 criterion; and there would be \$0 (no) reimbursement for PCI when a patient meets none of the criteria.

To those who would argue that such a scheme is draconian, Diamond and Kaul respond that it is “typical of social contracts: the incentives for obeying the rules are small; the disincentives for breaking them are large.”<sup>21</sup> The system would be particularly advantageous, they argue, because both the penalties for poor performance and the benefits of optimal performance would be immediate and obvious (i.e., lower vs. higher reimbursement), unlike the typical P4P scheme. “The importance of this [feature] should be self-evident to any parent,” Diamond and Kaul observe. “Just try to modify behavior with the hollow promise of relatively small rewards delayed long into the future.”<sup>21</sup>

Although the numerous administrative and contractual hurdles to the system proposed by Diamond and Kaul are obvious, support for the general principle that financial salience is important is found in Meterko et al.’s multistate survey (conducted in 2004) of primary care physicians participating in the Rewarding Results program, a national demonstration project on the effects of quality targets and financial incentives (response rate 32%).<sup>23</sup> Meterko et al. hypothesized that provider attitudes toward financial incentives depend on 7 attitudinal factors: “(1) *awareness and understanding* of the incentive program, (2) *salience* of the financial incentives, (3) *clinical relevance* of the quality targets, (4) *control* over the resources needed to achieve the quality targets, (5) *fairness* in the administration of the incentive program, (6) frequency and nature of *performance feedback* provided, and (7) possible *unintended consequences* associated with the pursuit of the quality targets.” (emphasis in original).

In an ordinary least squares regression analysis in which the dependent variable was “the perceived impact of quality targets and incentives on clinical practice behavior,” Meterko et al. accounted for covariates including years since residency, number of patients, medical school faculty appointment, specialty, and overall satisfaction with practice, then added attitudinal subscales (i.e., combinations of items that measured similar concepts, derived from an exploratory factor analysis and multitrait analysis) to an equation containing the covariates using a forward stepwise selection process. Meterko et al. found that the financial salience subscale, consisting of the views that the financial incentive “represents an opportunity for me to appreciably increase my income” and “is sufficiently large to compensate for expenditures that might be necessary in order to meet the quality target,” entered the equation first (i.e., it was the strongest predictor), explaining 13% of the variance in perceived impact of the incentive system. The second strongest predictor, explaining 5%, was cooperation: “I am able to get the cooperation of other physicians as needed to obtain this finan-

cial incentive” and “I am able to get the cooperation of support staff as needed to obtain this financial incentive.” Third, explaining 2%, was relevance, that is, whether the incentive was “good for my patients,” “based on sound medical science,” and “tied to a quality target that is clinically meaningful.”<sup>23</sup>

### Paying for Improvement Versus Performance

Greenberg et al. (2010) observed that quality improvement should be measured longitudinally for each medical practice rather than making comparisons among physicians—that is, rewarding for performance *improvement* rather than overall performance *level*.<sup>24</sup> This recommendation would seem to be supported by a naturalistic evaluation reported by Rosenthal et al. (2005), who compared the performance of 42 medical groups in the Pacific Northwest without financial incentives with 163 medical groups in California that had sufficient HMO membership (at least 1,000 commercial members and at least 100 Medicare Advantage members) to be included in the HMO’s quality incentive program (QIP).<sup>25</sup> The medical groups in the QIP were eligible for quarterly bonus payments of \$0.23 per member per month (PMPM) for each performance target that was met or exceeded, and the HMO paid out \$3.4 million to 97 of the 163 eligible physician groups (60%) that attained at least 1 quality performance target in the first year through April 2004. Only 1 of the 3 quality measures studied by Rosenthal et al. showed greater improvement in the difference-in-difference analysis for the QIP medical groups compared with the medical groups without a financial incentive: cervical cancer screening rates increased by 5.3 percentage points in the QIP groups versus 1.7 points in the groups without a financial incentive ( $P=0.02$ ). There was no difference in the performance change between the QIP and non-QIP medical groups for the other 2 measures, rates of mammography screening and A1c testing in patients with diabetes.<sup>25</sup>

Like the analysis of the HQID by Lindenauer et al., the analysis by Rosenthal et al. showed that medical groups with the lowest baseline performance improved the most (e.g., improvement of 6.6% in mammography screening) while higher-performing groups at baseline improved the least (e.g., improvement of 0.7% in mammography screening); yet, the groups with the highest baseline performance—those that improved the least—garnered 75% of the bonus payments. The findings by Rosenthal et al. suggest that paying physicians to reach fixed performance targets may result in small overall improvement in quality measures and spending most of the bonus payments on medical groups that would have performed exceptionally well anyway. This conclusion would appear to support the recommendation by Greenberg et al. that medical practices should be paid for incremental improvement rather than attainment of a fixed performance target. However, rewarding improvement rather than attainment might be viewed as condoning low performance.<sup>25</sup>

### Limitations of Available Evidence: A Caveat Emptor for MCOs

In considering the evidence for and against P4P as a tool to improve quality of care, it is perhaps helpful to consider the history of the promotion of disease management as a tool to simultaneously improve health care quality and reduce costs. In the trajectory of evidence-building that led first to over-hyped promises for disease management, then to a realization that disease management could not deliver on the hype, it was CMS that created a turning point in the debate with the funding of the randomized Medicare Health Support Experiment.<sup>4</sup> Whether CMS will lead the way again with a rigorous evaluation of P4P remains to be seen. Yet, a randomized trial of P4P is needed, not only because the results of observational comparative studies demonstrate that some degree of care improvement occurs regardless of financial incentives, but also because providers choosing to participate in P4P programs may systematically differ from those who decline participation. The possibility that selection bias and confounding effects have substantially compromised the currently available evidence about P4P looms large.

There is also a lack of information about how physicians will respond to quality initiatives in “real-life” practice as financial incentives become increasingly prevalent and therefore increasingly important. Greenberg et al. observed that many critical questions about how incentive programs will work are largely unaddressed, including care coordination (e.g., whether a patient’s LDL-C is the responsibility of the cardiologist, the primary care physician, or both); choice of metric (e.g., whether a physician should be able to choose from among clinically valid metrics in accepting a P4P arrangement); and infrastructure (e.g., whether the data used to document performance on the metric are available electronically or must be gathered using a cumbersome manual process).<sup>24</sup> Additional troubling possibilities are “cherry-picking” (i.e., the scuttling of higher-complexity cases to avoid financial penalties, especially in the face of inadequate severity adjustment)<sup>24</sup> or improvement in documentation only without actual care improvement, as Butler et al. (2006) found in a study of the effect of computerized prescription order entry on compliance with recommended medication therapy following hospitalization for AMI.<sup>26</sup>

If the disease management experience has anything relevant to teach payers about P4P, perhaps it is the danger of spending a lot of time and money chasing after small and clinically insignificant gains. To avoid that danger, payers need better information both about the outcomes of P4P and about providers’ perspectives on what interventions will work to improve quality in real-life clinical practice. So, in addition to randomized trials of financial incentives, an additional important part of the best strategy for MCOs may be to “ask your doctors if P4P is right for you.”

### Authors

KATHLEEN A. FAIRMAN, MA, is Associate Editor and Senior Methodology Reviewer, and FREDERIC R. CURTISS, PhD, RPh, CEBS, is Editor-in-Chief of the *Journal of Managed Care Pharmacy*.

AUTHOR CORRESPONDENCE: Kathleen A. Fairman, MA, Kathleen Fairman LTD, P.O. Box 31278, Phoenix, AZ 85046. Tel.: 602.867.1343; E-mail: kfairman@amcp.org.

### DISCLOSURES

The authors report no conflicts of interest related to the subjects or products discussed in this article.

### REFERENCES

1. Diamond GA, Kaul S. Evidence-based financial incentives for healthcare reform: putting it together. *Circ Cardiovasc Qual Outcomes*. 2009;2:134-40. Available at: <http://circoutcomes.ahajournals.org/cgi/reprint/2/2/134>. Accessed May 20, 2010.
2. Peterson LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med*. 2006;145(4):265-72. Available at: <http://www.annals.org/content/145/4/265.full.pdf+html>. Accessed May 20, 2010.
3. Berwick DM, DeParle N-A, Eddy DM, et al. Paying for performance: Medicare should lead (letter). *Health Aff (Millwood)*. 2003;22(6):8-10. Available at: <http://content.healthaffairs.org/cgi/reprint/22/6/8>. Accessed May 20, 2010.
4. Fairman KA, Curtiss FR. Still looking for health outcomes in all the wrong places? Misinterpreted observational evidence, medication adherence promotion, and value-based insurance design. *J Manag Care Pharm*. 2009;15(6):501-07. Available at: <http://www.amcp.org/data/jmcp/501-507.pdf>.
5. Brackbill ML, Kline VT, Sytsma CS, Call JT. Intervention to increase the proportion of acute myocardial infarction or coronary artery bypass graft patients receiving an order for aspirin at hospital discharge. *J Manag Care Pharm*. 2010;16(5):329-36. Available at: <http://www.amcp.org/data/jmcp/329-336.pdf>.
6. Centers for Medicare & Medicaid Services. QMIS measure details: AMI-2. Available at: [http://www.cms.gov/apps/QMIS/measure\\_details.asp?id=613](http://www.cms.gov/apps/QMIS/measure_details.asp?id=613). Accessed May 20, 2010.
7. Centers for Medicare & Medicaid Services. Reporting Hospital Quality Data for Annual Payment Update. Available at: [http://www.cms.gov/HospitalQualityInits/08\\_HospitalRHQDAPU.asp](http://www.cms.gov/HospitalQualityInits/08_HospitalRHQDAPU.asp). Accessed May 20, 2010.
8. Centers for Medicare & Medicaid Services. Hospital Quality Initiative overview. Available at: <http://www.cms.gov/HospitalQualityInits/>. Accessed May 21, 2010.
9. Centers for Medicare and Medicaid Services. Premier Hospital Quality Incentive Demonstration: rewarding superior quality care. Fact sheet. July 2009. Available at: <http://www.cms.gov/HospitalQualityInits/downloads/HospitalPremierFactSheet200907.pdf>. Accessed May 21, 2010.
10. Endsley S, Baker G, Kershner BA, Curtin K. What family physicians need to know about pay-for-performance. *Fam Pract Manag*. 2006;13(7):69-74. Available at: <http://www.aafp.org/fpm/2006/0700/p69.html>. Accessed May 22, 2010.
11. Centers for Medicare & Medicaid Services. Medicare News. Press release: Medicare demonstrations show paying for quality health care pays off. August 17, 2009. Available at: <http://www.cms.gov/HospitalQualityInits/downloads/HospitalPremierPressReleases20090817.pdf>. Accessed May 21, 2010.

12. Lester H, Schmittiel J, Selby J, et al. The impact of removing financial incentives from clinical quality indicators: longitudinal analysis of four Kaiser Permanente indicators. *BMJ*. 2010;340:c1898. Available at: [http://www.bmj.com/cgi/reprint/340/may11\\_1/c1898](http://www.bmj.com/cgi/reprint/340/may11_1/c1898). Accessed May 28, 2010.
13. Chung S, Palaniappan LP, Trujillo LM, Rubin HR, Luft HS. Effect of physician-specific pay-for-performance incentives in a large group practice. *Am J Manag Care*. 2010;16(2):e35-e42. Available at: [http://www.ajmc.com/media/pdf/AJMC\\_10feb\\_ChungXclsv\\_e35to42.pdf](http://www.ajmc.com/media/pdf/AJMC_10feb_ChungXclsv_e35to42.pdf). Accessed May 29, 2010.
14. Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. Quality of primary care in England with the introduction of pay for performance. *N Engl J Med*. 2007;357(2):181-90. Available at: <http://content.nejm.org/cgi/reprint/357/2/181.pdf>. Accessed May 29, 2010.
15. Glickman SW, Ou F-S, DeLong ER, et al. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA*. 2007;297(21):2373-80. Available at: <http://jama.ama-assn.org/cgi/reprint/297/21/2373>. Accessed May 19, 2010.
16. Lindenauer PK, Remus D, Roman S, et al. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med*. 2007;356(5):486-96. Available at: <http://content.nejm.org/cgi/reprint/356/5/486.pdf>. Accessed May 20, 2010.
17. Chen JY, Tian H, Juarez DT, et al. The effect of a PPO pay-for-performance program on patients with diabetes. *Am J Manag Care*. 2010;16(1):e11-e19. Available at: [http://www.ajmc.com/issue/managed-care/2010/2010-01-vol16-n01/AJMC\\_2010janChenWEB\\_e11t\\_e19](http://www.ajmc.com/issue/managed-care/2010/2010-01-vol16-n01/AJMC_2010janChenWEB_e11t_e19). Accessed May 21, 2010.
18. An LC, Bluhm JH, Foldes SS, et al. A randomized trial of a pay-for-performance program targeting clinician referral to a state tobacco quitline. *Arch Intern Med*. 2008;168(18):1993-99.
19. 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):206-12.
20. 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):1699-701.
21. Chung S, Palaniappan L, Wong E, Rubin H, Luft H. Does the frequency of pay-for-performance payment matter? Experience from a randomized trial. *Health Serv Res*. 2010;45(2):553-64.
22. Boden WE, O'Rourke RA, Teo KK, et al. Optimal medical therapy with or without PCI for stable coronary disease. *N Engl J Med*. 2007;356(15):1503-16. Available at: <http://content.nejm.org/cgi/reprint/356/15/1503.pdf>. Accessed May 20, 2010.
23. Meterko M, Young GJ, White B, et al. Provider attitudes toward pay-for-performance programs: development and validation of a measurement instrument. *Health Serv Res*. 2006;41(5):1959-78. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1955299/pdf/hesr0041-1959.pdf>. Accessed May 21, 2010.
24. Greenberg JO, Dudley JC, Ferris TG. Engaging specialists in performance-incentive programs. *N Engl J Med*. 2010;362(17):1558-60.
25. Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance from concept to practice. *JAMA*. 2005; 294(14):1788-93. Available at: <http://jama.ama-assn.org/cgi/reprint/294/14/1788>. Accessed May 29, 2010.
26. Butler J, Speroff T, Arbogast PG, et al. Improved compliance with quality measures at hospital discharge with a computerized physician order entry system. *Am Heart J*. 2006;151(3):643-53.