

The Effects of Pay-for-Performance Programs on Health, Health Care Use, and Processes of Care

A Systematic Review

Aaron Mendelson, BA; Karli Kondo, PhD; Cheryl Damberg, PhD; Allison Low, BA; Makalapua Motúapuaka, BA; Michele Freeman, MPH; Maya O'Neil, PhD; Rose Relevo, MLIS, MS; and Devan Kansagara, MD, MCR

Background: The benefits of pay-for-performance (P4P) programs are uncertain.

Purpose: To update and expand a prior review examining the effects of P4P programs targeted at the physician, group, managerial, or institutional level on process-of-care and patient outcomes in ambulatory and inpatient settings.

Data Sources: PubMed from June 2007 to October 2016; MEDLINE, PsycINFO, CINAHL, Business Economics and Theory, Business Source Elite, Scopus, Faculty of 1000, and Gartner Research from June 2007 to February 2016.

Study Selection: Trials and observational studies in ambulatory and inpatient settings reporting process-of-care, health, or utilization outcomes.

Data Extraction: Two investigators extracted data, assessed study quality, and graded the strength of the evidence.

Data Synthesis: Among 69 studies, 58 were in ambulatory settings, 52 reported process-of-care outcomes, and 38 reported patient outcomes. Low-strength evidence suggested that P4P programs in ambulatory settings may improve process-of-care outcomes over the short term (2 to 3 years), whereas data on

longer-term effects were limited. Many of the positive studies were conducted in the United Kingdom, where incentives were larger than in the United States. The largest improvements were seen in areas where baseline performance was poor. There was no consistent effect of P4P on intermediate health outcomes (low-strength evidence) and insufficient evidence to characterize any effect on patient health outcomes. In the hospital setting, there was low-strength evidence that P4P had little or no effect on patient health outcomes and a positive effect on reducing hospital readmissions.

Limitation: Few methodologically rigorous studies; heterogeneous population and program characteristics and incentive targets.

Conclusion: Pay-for-performance programs may be associated with improved processes of care in ambulatory settings, but consistently positive associations with improved health outcomes have not been demonstrated in any setting.

Primary Funding Source: U.S. Department of Veterans Affairs.

Ann Intern Med. 2017;166:341-353. doi:10.7326/M16-1881

Annals.org

For author affiliations, see end of text.

This article was published at Annals.org on 10 January 2017.

Pay-for-performance (P4P) programs provide financial rewards or penalties to individual health care providers, groups of providers, or institutions according to their performance on measures of quality. In theory, if properly targeted and designed, P4P programs would help drive the behavior of providers and health care systems to improve the quality of care delivered, reduce unnecessary use of expensive health care services, and improve patient health outcomes (1). The idea is particularly relevant in the United States, where serious and broad gaps in health care quality have been tied in part to the long-standing fee-for-service system, which may provide incentives for service volume rather than quality (2).

Despite their intuitive appeal, the promise of P4P programs in improving outcomes has not been empirically realized in past studies (3-6). The most recent systematic review examining the effectiveness of P4P programs in the United States found mixed evidence that P4P was associated with modest improvements in process-of-care outcomes but had little effect on patient outcomes (7). However, the literature has grown considerably since this review (which searched through 2012), and other countries, such as the United Kingdom, have gained considerable experience with large P4P initiatives that may provide information relevant to the United States. The purpose of the current review is

to update and expand the prior systematic review in order to summarize current understanding of the effects of P4P programs targeted at physicians, groups, and institutions on process-of-care and patient outcomes in ambulatory and outpatient settings in and outside the United States.

METHODS

This review was conducted according to a protocol that was developed using established reporting standards and posted to a public Web site (8) before the study was initiated (Appendix 1 of the Supplement, available at Annals.org). We used an analytic framework based on work by Damberg and colleagues (7) (Appendix 2 of the Supplement).

Data Sources and Searches

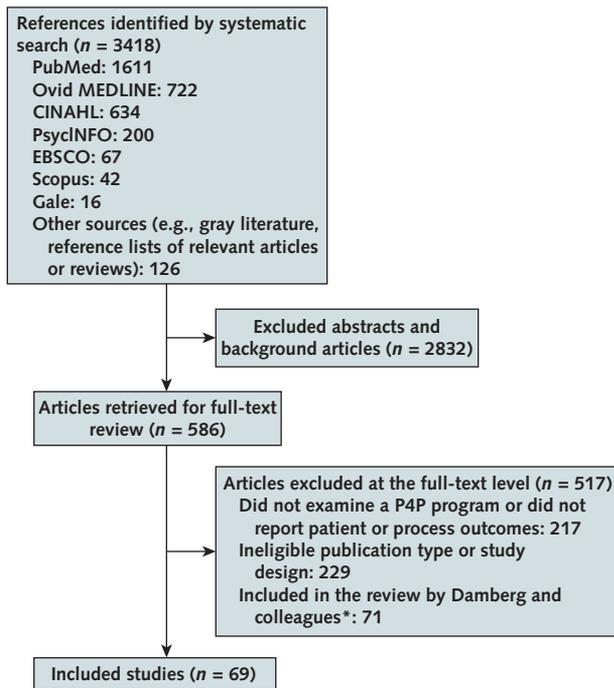
We searched the following databases for studies that evaluated P4P programs: PubMed (1 June 2007 to

See also:

Editorial comment 368

Web-Only

Supplement

Figure. Literature flow diagram.

P4P = pay-for-performance.

* The current systematic review updates and expands on the review by Damberg and colleagues (7).

6 October 2016), MEDLINE, PsycINFO, CINAHL, Business Economics and Theory, Business Source Elite, Scopus, Faculty of 1000, and Gartner Research (1 June 2007 to 29 February 2016). We also performed targeted Google and PubMed searches aimed at well-known P4P demonstrations. We obtained additional articles from reference lists of pertinent studies, reviews, editorials, and expert recommendations. The search strategies are detailed in **Appendix 3** of the **Supplement**.

Study Selection

Investigators reviewed titles and abstracts identified from literature searches. Two investigators independently assessed each potentially relevant article for inclusion using preestablished criteria (**Appendices 4 and 5** of the **Supplement**). We included English-language studies of adult patients that evaluated ambulatory care- or hospital-based P4P programs targeting health care providers at the individual, group, managerial, or institutional level and that reported any process-of-care, utilization, health, or intermediate health (clinical measures, such as a laboratory value or blood pressure) outcome. We included studies from other countries that have health systems similar to portions of the U.S. health care system. Studies examining only patient-targeted financial incentives, as well as payment models other than direct P4P, such as managed care, capitation, bundled payments, and accountable care organizations, were excluded. We also excluded studies that were not conducted in hospital or

ambulatory settings, such as studies in long-term care facilities or nursing homes.

We included clinical or cluster randomized, controlled trials (RCTs) of any size. We used a best-evidence approach, which is a method of specifying minimum inclusion criteria for nonrandomized studies (9). Inclusion of observational studies was limited to those with a comparison group, interrupted time series (ITS) studies, or large ($n > 10\,000$) cross-sectional or uncontrolled before-after studies. We excluded smaller uncontrolled studies because we had identified a large number of potentially relevant studies during a preliminary search and because the smaller uncontrolled studies were less likely to provide broadly applicable information given their limited scope and inherent methodological deficiencies.

Data Extraction and Quality Assessment

One investigator abstracted data elements from each included study, which were reviewed for accuracy by at least 1 additional investigator. We abstracted information on study design, sample size, country, program description, incentive structure (size and timing), target of the incentive, comparator, and outcomes (grouped as health, intermediate health, process-of-care, and utilization measures). **Appendices 6 and 7** of the **Supplement** report these data. We classified studies according to 4 broad groupings: RCTs, ITS studies, controlled before-after studies, and uncontrolled before-after studies. Two investigators independently assessed study quality using the Newcastle-Ottawa Scale (10) for observational studies and the Cochrane Risk-of-Bias tool (11) for RCTs (**Appendix 8** of the **Supplement**). Disagreements were resolved by consensus.

Data Synthesis and Analysis

We qualitatively synthesized the results of ambulatory and hospital studies separately and report process-of-care and patient outcomes for each setting. We synthesized results for specific P4P programs whenever possible. The review team evaluated the strength of the evidence according to guidance from the Agency for Healthcare Research and Quality (12). We did not perform meta-analysis because of the marked clinical heterogeneity across studies and the large number of observational studies.

Role of the Funding Source

The U.S. Department of Veterans Affairs Quality Enhancement Research Initiative supported this review but had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

RESULTS

Search Results

We reviewed 3418 titles and abstracts, identified 586 potentially eligible full-text articles, and ultimately included 69 studies (**Figure**). Fifty-eight studies were in

ambulatory settings (Table 1 and Appendix 6 of the Supplement), 11 were in hospital settings (Table 2 and Appendix 7 of the Supplement), 52 reported process-of-care outcomes, and 38 assessed patient outcomes. The studies examined a wide range of P4P programs with varying incentive structures, goals, and contexts. The programs also differed in their purposes and targets, but the largest number of studies focused on managing chronic conditions in the primary care setting. Studies were conducted in a wide range of countries, including the United Kingdom (27 studies), the United States (17 studies), Taiwan (13 studies), France (3 studies), the Netherlands (3 studies), Canada (3 studies), Australia (1 study), South Korea (1 study), and Italy (1 study). There were 2 RCTs and 67 observational studies (10 ITS studies, 37 controlled before-after studies, and 20 large uncontrolled before-after studies).

A large number of studies evaluated different aspects of 2 large-scale national programs: the United Kingdom's Quality and Outcomes Framework (QOF) (24 studies) and Taiwan's diabetes mellitus (DM-P4P) program (9 studies). The QOF is a nationwide program that began in 2004. It incentivizes primary care practices to achieve quality indicators that support clinical care and public health goals. Incentive payments can comprise up to approximately 30% of total income. Practices are aided by integrated health information technology that delivers automated prompts and decision support (36, 83). Taiwan's DM-P4P program, implemented in 2001, allows physicians to voluntarily enroll in the program, and they in turn are given freedom to choose which patients to enroll (51). From 2001 to 2006, incentives targeted process-of-care outcomes, which were augmented with intermediate health outcome measures after 2006.

Ambulatory Care–Based Programs

Process-of-Care Outcomes

We found 9 studies from the United States evaluating the effects of P4P on process-of-care outcomes (14, 16–20, 22–24). Most of these studies examined outcomes over 4 years and had an average follow-up of 2.5 years; very few studies reported longer-term data. One RCT found that individual incentives increased appropriate response to high blood pressure but not use of guideline-recommended antihypertensive medication (14). Of the 6 studies that reported positive results (16, 18, 19, 22–24), 1 did not have a control group (24), and selection bias was a serious concern in 3 others because of the way the control group was chosen (18, 22, 23). Two methodologically sound controlled before-after studies found no improvements in processes of care (17, 20).

In general, there was evidence across 17 studies in the United Kingdom (26–31, 33, 36–38, 41–47) that the QOF was associated with improvements in process-of-care measures, although the evidence was mixed among the more methodologically rigorous studies. There were 6 ITS studies. One showed substantial improvements in the prescription of long-acting reversible contraceptives (26), and another showed modest

improvement in the initiation of diabetes medications (27). Another study found increased rates of depression screening and diagnoses, but antidepressant prescribing remained unchanged (31). In the other 3 studies, improvements had begun well before QOF implementation, and postintervention trends did not show substantial improvement and, in fact, showed slower or decreased improvement over time (28–30).

Although many studies of Taiwan's DM-P4P program showed improvement in process-of-care measures, selection bias was a major concern (51–54, 58). Physicians voluntarily enrolled and were given discretion over which patients to enroll. Because the program lacked risk adjustment and, initially, a mechanism to disenroll patients, physicians had a strong incentive to enroll healthier patients (51). Indeed, enrolled patients were much healthier than nonenrolled patients. Moreover, at participating institutions, the pool of nonenrolled patients became sicker over time, indicating that healthier patients were being removed to participate in DM-P4P. Though many studies attempted to adjust for differences in the 2 groups by using propensity score matching, residual confounding was still an important potential issue given the many unmeasured factors that were likely to be related to enrollment decision making.

We found 13 non-U.S. studies that were not part of a larger P4P evaluation. Two of these studies were methodologically sound observational studies from Canada that reported contradictory results on screening and preventive measures (66, 67). An ITS study found modest increases in colorectal cancer screening but no effects on cervical and breast cancer screening (66). However, a controlled before-after study found modest increases for colorectal cancer screening, mammography, flu shots, and Papanicolaou smears (67). It was difficult to draw strong conclusions from the other 11 studies because of disparities in the programs' targets and designs and the study settings, as well as the low quality of the study designs (49, 50, 61–65, 68–71).

Patient Outcomes

Health Outcomes. Ten studies evaluated health outcomes in ambulatory settings (39, 44, 51, 52, 55–57, 59–61). Eight of the studies (most of which found positive results) were conducted in Taiwan and should be interpreted with caution due to selection bias, as described earlier (51, 52, 55–57, 59–61). Two large uncontrolled before-after studies of QOF reported no improvements in health outcomes (39, 44). One assessed the correlation among regional QOF performance, all-cause mortality, and condition-specific mortality (39). It found that better performance on both the aggregate of QOF quality indicators and a subset of intermediate outcome indicators did not correlate with reduced mortality. Another study found that chronic obstructive pulmonary disease (COPD) prevalence actually increased from 1.27% to 1.45% after QOF implementation (44). Given the time needed to develop COPD and

Table 1. Findings From Studies of Ambulatory-Based Pay-for-Performance Programs

Study, Year (Reference)	Sample Size	Target of the Intervention	Size of the Incentive	Timing	Findings				
					Process-of-Care Outcomes	Patient Health Outcomes	Patient Utilization	Patient Intermediate Outcomes	
United States									
RCTs									
Asch et al, 2015 (13)	340 primary care physicians 1503 patients	Provider	\$256 per quarter per patient	Semiannual	-	-	-	Null	
Petersen et al, 2013 (14) Petersen et al, 2016 (15)*	83 physicians	Group and provider	\$9.10 per measure, \$4270 average group total, and \$1648 average per provider	Every 4 mo	Mixed	-	-	Positive	
Controlled before-after studies									
Esse et al, 2013 (16)	4240 patients	NR	NR	NR	Positive	-	Null	-	
Friedberg et al, 2014 (17)	61 practices 120 202 patients	Group	\$20 000 lump sum in year 1 and annual bonus payments based on clinic size and National Committee for Quality Assurance medical home recognition level	Annual	Null	-	Null	-	
Kruse et al, 2013 (18)	20 774 patients	Group	Withheld 3%–4.8% of practice revenue	NR	Positive	-	-	-	
Lemak et al, 2015 (19)	3.2 million patients	Group and provider	10% increase in certain FFS activities	Biannual	Positive	-	-	-	
Rosenthal et al, 2016 (20)	3 state Medicaid programs	Group and provider	Either 1-time payment of \$200 with additional payments from \$17–\$60, \$100 per patient, or tiered case-management fee from \$0.10–\$0.85	NR	Null	-	Mixed	-	
Rosenthal et al, 2016 (21)	98 000 patients	Group	NR	Annual	-	-	Mixed	-	
Share and Mason, 2012 (22)	994 practices	Groups	Up to 4.7% enhanced FFS	Annual	Positive	-	Positive	-	
Young et al, 2012 (23)	171 physicians	Provider	Up to \$15 000	NR	Positive	-	-	-	
Uncontrolled before-after study									
Torchiana et al, 2013 (24)	1700 physicians	Providers	2% of annual income	6 mo	Positive	-	Positive	-	
United Kingdom: QOF									
ITS studies									
Alshamsan et al, 2012 (25)	7434 patients 29 family practices	Group	25%–30% of physician income	Annual	-	-	-	Mixed	
Arrowsmith et al, 2014 (26)	581 practices	-	-	-	Positive	-	-	-	
Gallagher et al, 2015 (27)	516 primary care practices	-	-	-	Positive	-	-	-	
Kendrick et al, 2015 (28)	191 117 patients	-	-	-	Mixed	-	-	-	
Kontopantelis et al, 2013 (29)	23 920 patients	-	-	-	Positive	-	-	-	
MacBride-Stewart et al, 2008 (30)	92 practices	-	-	-	Negative	-	-	-	
McLintock et al, 2014 (31)	65 general practices	-	-	-	Mixed	-	-	-	
Vamos et al, 2011 (32)	154 945 patients	-	-	-	-	-	-	Mixed	

Continued on following page

Table 1—Continued

Study, Year (Reference)	Sample Size	Target of the Intervention	Size of the Incentive	Timing	Findings			
					Process- of-Care Outcomes	Patient Health Outcomes	Patient Utilization	Patient Intermediate Outcomes
Controlled before- after studies								
Doran et al, 2011 (33)	148 practices 653 500 patients	-	-	-	Positive	-	-	-
Harrison et al, 2014 (34)	6975 practices	-	-	-	-	-	Positive	-
Karunaratne et al, 2013 (35)	10 040 patients	-	-	-	-	-	-	Positive
Sutton et al, 2010 (36)	315 providers	-	-	-	Positive	-	-	-
Uncontrolled before- after studies								
Calvert et al, 2009 (37)	147 practices	-	-	-	Positive	-	-	Mixed
Hamilton et al, 2016 (38)	41 239 patients	-	-	-	Positive	-	-	Positive
Kontopantelis et al, 2015 (39)	8647 general practices	-	-	-	-	Null	-	-
Millett et al, 2009 (40)	422 practices 154 945 patients	-	-	-	-	-	-	Mixed
Murray et al, 2010 (41)	3200 patients	-	-	-	Positive	-	-	Positive
Norbury et al, 2011 (42)	315 practices 300 000 patients	-	-	-	Positive	-	-	-
Simpson et al, 2011 (43)	315 practices	-	-	-	Positive	-	-	Positive
Smith et al, 2008 (44)	2 020 424 patients	-	-	-	Positive	Negative	-	-
Szatkowski et al, 2011 (45)	2 million patients	-	-	-	Positive	-	-	-
Taggar et al, 2012 (46)	2 million patients	-	-	-	Positive	-	-	-
Tahrani et al, 2007 (47)	66 practices 460 000 patients	-	-	-	Positive	-	-	Positive
Vaghela et al, 2009 (48)	8192 practices	-	-	-	-	-	-	Positive
United Kingdom: other programs								
Controlled before- after studies								
Mason et al, 2015 (49)	346 300 patients	Provider	NR	NR	Negative	-	-	-
Kalwij et al, 2012 (50)	95 general practices	Group	Either based on proportion of eligible screens (£100–£2600) or £6–£15 per screen	Annual	Positive	-	-	-
Taiwan: DM-P4P								
Controlled before- after studies								
Chang et al, 2012 (51)	699 876 patients	Provider	\$151–\$181 per patient	Annual	Positive	Positive	-	-
Chen et al, 2016 (52)	2090 patients	-	-	-	Null	Null	Null	-
Chen and Cheng, 2016 (53)	8351 patients	-	-	-	Positive	-	Positive	-
Cheng et al, 2012 (54)	140 000 patients 3582 providers	-	-	-	Positive	-	Mixed	-
Hsieh et al, 2015 (55)	74 529 patients	-	-	-	-	Positive	-	-
Hsieh et al, 2016 (56)	34 710 patients	-	-	-	-	Positive	-	-
Liao et al, 2016 (57)	32 084 patients	-	-	-	-	Positive	-	-

Continued on following page

Table 1—Continued

Study, Year (Reference)	Sample Size	Target of the Intervention	Size of the Incentive	Timing	Findings			
					Process-of-Care Outcomes	Patient Health Outcomes	Patient Utilization	Patient Intermediate Outcomes
Lee et al, 2010 (58)	38 671 patients	-	-	-	Positive	-	Positive	-
Tan et al, 2014 (59)	260 patients	-	-	-	-	Positive	-	-
Taiwan: other								
Controlled before-after studies								
Lee et al, 2015 (60)	6009 patients	Group	Enhanced FFS, \$16 case finding fee, and \$33-\$66 treatment completion bonus	NR	-	Positive	Null	-
Li et al, 2010 (61)	33 000 patients	-	Enhanced FFS, \$16 case finding fee, and \$33-\$66 treatment completion bonus	NR	Null	Positive	-	-
Chen et al, 2016 (62)	21 643 patients	Physician	Up to \$52 per patient	Annual	Positive	-	-	-
The Netherlands								
Controlled before-after studies								
Martens et al, 2007 (63)	237 physicians	Provider	NR	NR	Null	-	-	-
Pechlivanoglou et al, 2015 (64)	169 000 patients	Provider	€0.25-€0.75 per patient	NR	Null	-	-	-
Uncontrolled before-after study								
Kirschner et al, 2013 (65)	65 practices	Group	Up to 10% of practice income	4 mo after data collection	Mixed	-	-	-
Canada								
ITS study								
Kiran et al, 2014 (66)	4992 physicians	Provider	3% of gross income	NR	Mixed	-	-	-
Controlled before-after study								
Li et al, 2014 (67)	2154 physicians	Provider	Up to 10% of physician revenue	Annual	Positive	-	-	-
France								
Controlled before-after studies								
Michel-Lepage and Ventelou, 2016 (68)	4622 general practitioners	Provider	NR	Annual	Positive	-	-	-
Sicsic and Franc, 2016 (69)	16 428 physicians 50 742 women	Physician	Up to €5000	NR	Null	-	-	-
Uncontrolled before-after study								
Rat et al, 2014 (70)	1350 general practitioners	Provider	Up to €5000	NR	Mixed	-	-	-
Australia								
Uncontrolled before-after study								
Greene, 2013 (71)	541 general practitioners	Group	Signing bonus (\$250-\$1000), \$20-\$40 per patient	NR	Null	-	-	-

DM-P4P = diabetes mellitus pay-for-performance program; FFS = fee-for-service; ITS = interrupted time series; NR = not reported; QOF = Quality and Outcomes Framework; RCT = randomized, controlled trial.

* Substudy of original data focusing on black patients. The sample size consisted of 67 physicians, and the average total payment was \$2744.

that most QOF indicators focused on managing COPD rather than preventing it, the implications of these findings are unclear.

Studies with high risk of bias generally found positive effects associated with DM-P4P (51, 52, 55-57, 59) and the similarly structured tuberculosis P4P program (60, 61). However, given the limitations already highlighted, such results are difficult to interpret.

Utilization Outcomes. We found 6 studies from the United States (16, 17, 20-22, 24), 5 studies from Taiwan (52-54, 58, 60), and 1 QOF study (34) reporting utilization outcomes.

The 6 studies from the United States reported mixed findings on the effects of P4P on utilization, although studies with the strongest designs showed no effect. One rigorously controlled study examined a P4P intervention that provided bonuses to practices that achieved advanced medical home status and found no effect on all-cause hospitalizations, all-cause emergency department (ED) visits, or ambulatory care-sensitive ED visits (17). Ambulatory care-sensitive hospitalizations actually increased in the second year of the intervention. Another controlled before-after study examined P4P in 3 state Medicaid programs and found no changes in any of the states for ED visits and inconsistent findings on inpatient utilization (20). A study examining a P4P program in medical homes targeting improved diabetes screenings and care found reductions in ED use and primary care visits but not in 6 other utilization measures (21). One study of a Medicare Advantage plan that rewarded physicians for providing evidence-based care to patients with heart failure found no effect on acute admissions or ED visits (16). Two studies lacking appropriate control groups showed improvement in ED use (22, 24).

Studies in Taiwan generally found reductions in hospital use associated with P4P (52-54, 58, 60). Again, due to the high likelihood of selection bias, these studies should be interpreted with caution.

A QOF study found a sustained reduction in ambulatory care-sensitive ED admissions (34).

Intermediate Health Outcomes. Twelve studies reported 1 or more intermediate health outcomes (13, 14, 25, 32, 35, 37, 38, 40, 41, 43, 47, 48). There were 2 RCTs with low risk of bias conducted in the United States. One RCT ($n = 1503$) evaluated the effect of a P4P program on low-density lipoprotein cholesterol levels (13). Physicians were given monthly patient progress reports and were eligible for comparatively large P4P bonuses (\$256 quarterly per patient) that were separated from other funding sources to highlight their relevance. Physicians received average total incentive payments of \$3246. The difference in low-density lipoprotein cholesterol level between patients seen by physicians in the P4P and control groups was not significant (2.8 mg/dL [95% CI, -1.7 to 7.4 mg/dL]; $P = 0.66$).

The other RCT (14) was included in the prior review by Damberg and colleagues, but a substudy was recently published (15). The original trial compared the effect of financial incentives earned for controlled blood pressure or response to uncontrolled blood pressure across 4 groups: incentives directed to individual physicians, practices, or both, or no incentives

(14). The study included 77 physicians; payments and performance feedback were delivered to physicians at the end of each 4-month performance period. The average total payment for physicians completing the entire program was \$2744. A higher proportion of patients achieved one or both measures in the individual physician incentive group than the control group (difference, 8.36% [CI, 2.4% to 13.0%]; $P = 0.005$), although the differences were not significant in the other 2 intervention groups. The recently published substudy found that the proportion of patients achieving control was not significantly higher in the incentive group (15).

Ten observational studies examining QOF reported mixed findings on intermediate outcomes (25, 32, 35, 37, 38, 40, 41, 43, 47, 48), but methodologically stronger studies suggested that QOF had little effect. Uncontrolled studies suggested large improvements in blood pressure control, cholesterol levels, and hemoglobin A_{1c} (HbA_{1c}) control. However, higher-quality studies that accounted for time trends failed to replicate these findings (25, 32). One short-term ITS study found that blood pressure control and cholesterol levels improved but HbA_{1c} control worsened relative to the underlying trend (32). A longer-term ITS study found that although mean cholesterol and HbA_{1c} levels and blood pressure control had been improving before QOF implementation, only systolic blood pressure continued to improve afterward. Diastolic blood pressure, mean cholesterol levels, and HbA_{1c} levels actually worsened relative to the pre-QOF trend (25).

Hospital-Based Programs

Process-of-Care Outcomes

Eight studies examined process-of-care measures in the hospital setting (74-77, 79-82). Controlled before-after studies from the United States and Canada generally failed to find improvements in care processes (74, 75), although 1 study from Canada did report modest reductions in ED wait times (80). One controlled study from Taiwan found that P4P-enrolled patients with breast cancer received better-quality care than nonenrolled patients (79). Uncontrolled studies reported larger improvements (76, 77, 81, 82).

Patient Outcomes

Health Outcomes. Pay-for-performance programs generally did not decrease mortality or improve patient experience in 5 studies in hospital settings (73, 74, 78, 79, 82). High-quality studies examining the U.K. Hospital Quality Incentive demonstration and the U.S. Hospital Value-Based Purchasing (HVBP) programs did not find a link between mortality and targeted conditions (73, 78). One short-term controlled before-after study found no immediate change in patient experience associated with the HVBP program (74). One uncontrolled study found that mortality related to hemorrhagic strokes did not decrease after implementation of P4P (82). A study from Taiwan indicated that P4P patients had improved breast cancer survival (79).

Utilization Outcomes. One ITS study reported utilization outcomes (72) and found that hospital

Table 2. Findings From Studies of Hospital-Based Pay-for-Performance Programs

Study, Year (Reference)	Country	Sample Size	Target of the Intervention	Size of the Incentive	Timing	Findings				
						Process-of-Care Outcomes	Patient Health Outcomes	Patient Utilization	Patient Intermediate Outcomes	
United States										
ITS study										
Zuckerman et al, 2016 (72)	-	3387 hospitals	Group	1%-3% of diagnosis-related group payments	Annual	-	-	Positive	-	-
Controlled before-after studies										
Figueroa et al, 2016 (73)	-	2919 hospitals 2.25 million patients	Group	1%-2% Medicare payment	Annual	-	Null	-	-	-
Ryan et al, 2015 (74)	-	2873 hospitals	Group	1%-2% bonus	NR	Null	Null	-	-	-
Ryan et al, 2014 (75)	-	260 hospitals	Group	1% of hospital payments	NR	Null	-	-	-	-
Uncontrolled before-after studies										
Andriole et al, 2010 (76)	-	224 physicians	Providers	\$4000	Semiannual	Positive	-	-	-	-
Benzer et al, 2014 (77)	-	128 VA medical centers	Group	NR	NR	Positive	-	-	-	-
Other countries										
Controlled before-after studies										
Kristensen et al, 2014 (78)*	United Kingdom	161 hospitals	Group	Up to \$5 million bonus changed to up to \$5 million penalty	NR	-	Null	-	-	-
Kuo et al, 2011 (79)†	Taiwan	1393 patients	Group	Enhanced bundled payment with a bonus tied to survival	NR	Positive	Positive	-	-	-
Vermeulen et al, 2016 (80)	Canada	70 hospitals	Group	NR	NR	Mixed	-	-	-	-
Uncontrolled before-after studies										
Colais et al, 2013 (81)	Italy	12 433 patients	Group	Full diagnosis-related group for target attainment, reduced diagnosis-related group for missed target	NR	Positive	-	-	-	-
Yang et al, 2016 (82)	South Korea	201 hospitals	Group	0.5%-1.0% total reimbursement	NR	Positive	Null	-	-	-

ITS = interrupted time series; NR = not reported; VA = U.S. Department of Veterans Affairs.

* Hospital Quality Incentive Demonstration.

† Pay-for-performance program for breast cancer care.

readmissions among Medicare fee-for-service patients decreased sharply for approximately 2 years after implementation of the Hospital Readmissions Reduction Program; improvements continued thereafter but at a substantially lower rate. Although readmission reductions were seen for various conditions, they decreased more among the measures that were specifically targeted by the program than those that were not.

DISCUSSION

This systematic review of 69 studies updated and expanded on a previous review that had focused on U.S. programs and reported similar findings (7). The strength of the evidence and key results are summarized in Table 3. Overall, in the ambulatory setting, we found low-strength evidence that P4P programs may improve process-of-care outcomes over the short term

(2 to 3 years). Evidence on the longer-term effects of P4P programs was limited. Many of the studies reporting positive findings were conducted in the United Kingdom, where incentives were much larger than any P4P programs in the United States. The largest improvements were seen in areas where baseline performance was poor. We found low-strength evidence that P4P had little to no effect on intermediate health outcomes (changes in laboratory measures), though there were inconsistencies among study results. The evidence examining patient health outcomes was insufficient because few methodologically rigorous studies reported these outcomes. In the hospital setting, low-strength evidence showed that P4P had a neutral effect on patient health outcomes and a positive effect on reducing hospital readmissions.

Although many studies found positive effects associated with P4P programs, the results were inconsistent across studies, the magnitude of effect was often small, and it was difficult to confidently ascribe observed changes in outcomes to the intervention itself because of the observational nature of most studies and their specific methodological flaws. To better characterize the breadth of programs that have been evaluated, we included large uncontrolled studies reporting outcomes before and after program implementation. However, in all of these studies, the 2 measurements potentially reflect the peak and average of normally expected measurement variation (a phenomenon known as regression to the mean). The controlled before-after studies do not have this same issue, but the choice of control group was problematic in many studies because either the patients who qualified for a P4P program differed systematically from those who did not, or

the participating providers or practices differed substantially from those that did not participate. The ITS studies were useful because they accounted for trends in outcomes before the intervention. Indeed, several of these studies showed that improvements in outcomes had begun before P4P implementation. It is unclear whether these reflected secular trends in health care or practice changes in anticipation of intervention implementation.

Our findings complement and add to prior reviews, which have also generally found that P4P programs have not been consistently effective in improving patient outcomes (3-7). There are several reasons why this might be the case. First, especially in the era of modern health reform, P4P programs have been implemented and assessed in settings where other effective quality improvement interventions—such as public reporting, audit and feedback, and electronic decision-support tools—may have been deployed (84). The incremental benefit of P4P may therefore have been more difficult to demonstrate.

Second, it is possible that P4P programs have not tested the “best” incentive structures and payment mechanisms. Experts have suggested the importance of designing P4P programs using the principles of behavioral economics, in which such factors as payment size, timing, and frequency are believed to have important influences on individual behavior (85). In health care, we have not found strong empirical data to help determine the most successful incentive structure (86). It is interesting to consider the United Kingdom's QOF program, which accounted for nearly 40% of the included studies in our review, alongside U.S. efforts. Studies of QOF found that incentivized process-of-care

Table 3. Strength of the Evidence

Outcome Type	Study Design	Study Limitations	Consistency	Strength of Evidence	Summary of Findings
Ambulatory					
Process	1 RCT 7 ITS studies 23 controlled before-after studies 13 uncontrolled before-after studies	Medium	Inconsistent	Low	Much of the evidence for positive effects comes from the QOF program. Little evidence of long-term effects; biggest improvements seen in areas with poor baseline performance.
Health	8 controlled before-after studies 2 uncontrolled before-after studies	High	Inconsistent	Insufficient	Most of the controlled studies have significant selection bias, and the 2 uncontrolled studies do not provide sufficient information to draw conclusions.
Utilization	11 controlled before-after studies 1 uncontrolled before-after study	Medium	Inconsistent	Low	Stronger study designs showed no effect.
Intermediate	2 RCTs 2 ITS studies 1 controlled before-after study 7 uncontrolled before-after studies	Medium	Inconsistent	Low	No consistently large effects; stronger observational studies showed no effect; 2 trials produced conflicting results.
Hospital					
Process	4 controlled before-after studies 4 uncontrolled before-after studies	High	Inconsistent	Low	Stronger study designs showed little to no effect.
Health	53 controlled before-after studies	Medium	Inconsistent	Low	The strongest studies showed no effect.
Utilization	1 ITS study	Medium	-	Low	1 national U.S. study showed a significant reduction in readmissions after introduction of a hospital-level financial penalty program.
Intermediate	1 controlled before-after study	High	-	Insufficient	1 study with short-term follow-up assessing patient experience.

ITS = interrupted time series; QOF = Quality and Outcomes Framework; RCT = randomized, controlled trial.

measures can lead to improvements, especially in the early years of program implementation, but the rate of improvement slowed over time and there was no clear evidence that QOF improved patient outcomes. Whereas the P4P programs in the United States tended to be implemented within health systems or payers and involve relatively small incentives, QOF is the largest P4P program ever attempted in health care. It was implemented nationally with a single payer that includes virtually all general practitioners and provides practices with up to 30% of their annual income.

Finally, P4P programs are very complex health system interventions that have been implemented in various ways. In a related article, we examined the implementation factors that might mediate the potentially beneficial and harmful effects of P4P programs (86). We systematically reviewed studies of implementation factors and also conducted interviews with experts in the field of P4P. Although direct evidence was inadequate to draw strong conclusions, we found that provider buy-in and alignment of measures with organizational goals were likely to be important in sustaining effective programs. We found that measures that were transparently developed from the evidence base and that were focused on improving clinical processes and patient outcomes rather than measures of efficiency were more likely to be effective. We also found that the overall number of incentives in place at any one time needs to be carefully considered. Given the evidence that the most substantial gains were consistently seen in areas of poor baseline performance, we suggested that organizations use incentives in the most-needed areas, review measures regularly, and discontinue them after achieving sustained improvements.

Our review has several important limitations. The evidence is limited by methodological flaws, variation in program and population characteristics, and limited reporting on secular trends in health care. We chose to include studies from other countries because the breadth of experience with P4P might be informative for some U.S. health systems, but we acknowledge that there are also limitations in applying findings from other countries broadly in the United States. Our review expands on a prior review, so it is possible we did not include some individual studies that are informative, though these probably would not have altered our summary findings.

The policy implications of our findings are open to interpretation. In the absence of strong evidence of benefit, it may be particularly important to consider the potential harms and costs associated with P4P. We recently published a systematic review of the unintended consequences of P4P: There was very limited evidence assessing the extent of gaming, no consistent evidence of a negative effect on health disparities, and a small amount of evidence suggesting the potential for both positive and negative effects on unincentivized measures (87). The costs and burden of documentation and reporting requirements associated with P4P programs are also important to consider but have not been studied extensively. Qualitative studies have found that

providers perceive P4P programs as imposing a considerable burden and threatening clinical autonomy (88–90). A recent survey study found that U.S. health care providers self-report spending about 15 hours per week reporting and interpreting data for measures, which translates into billions of dollars in opportunity cost (91). Indeed, the United Kingdom decided to scale back its QOF program after 10 years of experience, in part because of provider concerns and the inconsistency of data demonstrating long-term benefit (92).

On the other hand, P4P programs have likely been effective in some areas, most notably in improving processes of care. The lack of evidence on patient outcomes may reflect deficiencies in the methods that have been used to study these effects and the likelihood that it takes a long time for process-of-care improvements to translate into large-scale patient outcome improvements (93).

In summary, we found low-strength, contradictory evidence that P4P programs could improve processes of care, but we found no clear evidence to suggest that they improve patient outcomes. Value-based purchasing is a cornerstone of the coming Medicare reform known as the Medicare Access and CHIP Reauthorization Act, so P4P will remain a fixture in U.S. health care for the foreseeable future (94). Whether the inconsistency of positive findings suggests that P4P, broadly speaking, is unlikely to have large effects or is related to the marked differences in program design, patient population, and incentive target is unclear.

From VA Portland Health Care System, and Oregon Health & Science University, Portland, Oregon, and RAND Corporation, Santa Monica, California.

Disclaimer: The views and conclusions expressed in this article are those of the authors, who are responsible for its content, and do not necessarily represent the views of the U.S. Department of Veterans Affairs or the U.S. government. Therefore, no statement in this article should be construed as an official position of the U.S. Department of Veterans Affairs.

Financial Support: By the U.S. Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Quality Enhancement Research Initiative, Evidence-based Synthesis Program (project 05-225).

Disclosures: Authors have disclosed no conflicts of interest. Forms can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M16-1881.

Reproducible Research Statement: Study protocol and data set: See the Supplement. Statistical code: Not applicable.

Requests for Single Reprints: Devan Kansagara, MD, MCR, VA Portland Health Care System, Mail Code R&D 71, 3710 SW US Veterans Hospital Road, Portland, OR 97239-2999; e-mail, kansagar@ohsu.edu.

Current author addresses and author contributions are available at Annals.org.

References

- Epstein AM, Lee TH, Hamel MB. Paying physicians for high-quality care. *N Engl J Med*. 2004;350:406-10. [PMID: 14736934]
- Institute of Medicine. Report Brief. Crossing the Quality Chasm: A New Health System for the 21st Century. 2001. Accessed at <http://iom.nationalacademies.org/reports/2001/crossing-the-quality-chasm-a-new-health-system-for-the-21st-century.aspx> on 15 November 2016.
- Houle SK, McAlister FA, Jackevicius CA, Chuck AW, Tsuyuki RT. Does performance-based remuneration for individual health care practitioners affect patient care?: A systematic review. *Ann Intern Med*. 2012;157:889-99. [PMID: 23247940] doi:10.7326/0003-4819-157-12-201212180-00009
- Scott A, Sivey P, Ait Ouakrim D, Willenberg L, Naccarella L, Furler J, et al. The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrane Database Syst Rev*. 2011:CD008451. [PMID: 21901722] doi:10.1002/14651858.CD008451.pub2
- Eijkenaar F, Emmert M, Scheppach M, Schöffski O. Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy*. 2013;110:115-30. [PMID: 23380190] doi:10.1016/j.healthpol.2013.01.008
- Gillam SJ, Siriwardena AN, Steel N. Pay-for-performance in the United Kingdom: impact of the Quality and Outcomes Framework: a systematic review. *Ann Fam Med*. 2012;10:461-8. [PMID: 22966110] doi:10.1370/afm.1377
- Damberg CL, Sorbero ME, Lovejoy SL, Martsof GR, Raaen L, Mandel D. Measuring Success in Health Care Value-Based Purchasing Programs: Findings from an Environmental Scan, Literature Review, and Expert Panel Discussions. Santa Monica, CA: RAND Corporation; 2014. Accessed at www.rand.org/pubs/research_reports/RR306.html on 15 November 2016.
- Evidence-based Synthesis Program Coordinating Center. ESP Reports in Progress. Washington, DC: U.S. Department of Veterans Affairs; 2014. Accessed at www.hsrd.research.va.gov/publications/esp/in_progress.cfm on 20 August 2016.
- Treadwell JR, Singh S, Talati R, McPheeters ML, Reston JT. A Framework for "Best Evidence" Approaches in Systematic Reviews. Rockville, MD: Agency for Healthcare Research and Quality; 2011.
- Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of non-randomised studies in meta-analyses. Ottawa, Ontario, Canada: Ottawa Hospital Research Institute; 2014. Accessed at www.ohri.ca/programs/clinical_epidemiology/oxford.asp on 15 November 2016.
- Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. Cochrane Collaboration. 2011. Accessed at <http://handbook.cochrane.org> on 15 November 2016.
- Berkman N, Lohr K, Ansari M, McDonagh M, Balk E, Whitlock E, et al. Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update. *Methods Guide for Comparative Effectiveness Reviews*. AHRQ publication no. 13(14)-EHC130-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2013. Accessed at www.effectivehealthcare.ahrq.gov/ehc/products/322/998/MethodsGuideforCERs_Viswanathan_IndividualStudies.pdf on 15 November 2016.
- Asch DA, Troxel AB, Stewart WF, Sequist TD, Jones JB, Hirsch AG, et al. Effect of financial incentives to physicians, patients, or both on lipid levels: a randomized clinical trial. *JAMA*. 2015;314:1926-35. [PMID: 26547464] doi:10.1001/jama.2015.14850
- Petersen LA, Simpson K, Pietz K, Urech TH, Hysong SJ, Profit J, et al. Effects of individual physician-level and practice-level financial incentives on hypertension care: a randomized trial. *JAMA*. 2013;310:1042-50. [PMID: 24026599] doi:10.1001/jama.2013.276303
- Petersen LA, Ramos KS, Pietz K, Woodard LD. Impact of a pay-for-performance program on care for black patients with hypertension: important answers in the era of the Affordable Care Act. *Health Serv Res*. 2016. [PMID: 27329344] doi:10.1111/1475-6773.12517
- Esse T, Serna O, Chitnis A, Johnson M, Fernandez N. Quality compensation programs: are they worth all the hype? A comparison of outcomes within a Medicare Advantage heart failure population. *J Manag Care Pharm*. 2013;19:317-24. [PMID: 23627577]
- Friedberg MW, Schneider EC, Rosenthal MB, Volpp KG, Werner RM. Association between participation in a multipayer medical home intervention and changes in quality, utilization, and costs of care. *JAMA*. 2014;311:815-25. [PMID: 24570245] doi:10.1001/jama.2014.353
- Kruse GR, Chang Y, Kelley JH, Linder JA, Einbinder JS, Rigotti NA. Healthcare system effects of pay-for-performance for smoking status documentation. *Am J Manag Care*. 2013;19:554-61. [PMID: 23919419]
- Lemak CH, Nahra TA, Cohen GR, Erb ND, Paustian ML, Share D, et al. Michigan's fee-for-value physician incentive program reduces spending and improves quality in primary care. *Health Aff (Millwood)*. 2015;34:645-52. [PMID: 25847648] doi:10.1377/hlthaff.2014.0426
- Rosenthal MB, Landrum MB, Robbins JA, Schneider EC. Pay for performance in Medicaid: evidence from three natural experiments. *Health Serv Res*. 2016;51:1444-66. [PMID: 26708000] doi:10.1111/1475-6773.12426
- Rosenthal MB, Alidina S, Friedberg MW, Singer SJ, Eastman D, Li Z, et al. A difference-in-difference analysis of changes in quality, utilization and cost following the Colorado Multi-payer Patient-Centered Medical Home Pilot. *J Gen Intern Med*. 2016;31:289-96. [PMID: 26450279] doi:10.1007/s11606-015-3521-1
- Share DA, Mason MH. Michigan's Physician Group Incentive Program offers a regional model for incremental 'fee for value' payment reform. *Health Aff (Millwood)*. 2012;31:1993-2001. [PMID: 22949448] doi:10.1377/hlthaff.2012.0328
- Young GJ, Beckman H, Baker E. Financial incentives, professional values and performance: a study of pay-for-performance in a professional organization. *J Organ Behav*. 2012;33:964-83.
- Torchiana DF, Colton DG, Rao SK, Lenz SK, Meyer GS, Ferris TG. Massachusetts General Physicians Organization's quality incentive program produces encouraging results. *Health Aff (Millwood)*. 2013;32:1748-56. [PMID: 24101064] doi:10.1377/hlthaff.2013.0377
- Alshamsan R, Lee JT, Majeed A, Netuveli G, Millett C. Effect of a UK pay-for-performance program on ethnic disparities in diabetes outcomes: interrupted time series analysis. *Ann Fam Med*. 2012;10:228-34. [PMID: 22585887] doi:10.1370/afm.1335
- Arrowsmith ME, Majeed A, Lee JT, Saxena S. Impact of pay for performance on prescribing of long-acting reversible contraception in primary care: an interrupted time series study. *PLoS One*. 2014;9:e92205. [PMID: 24694949] doi:10.1371/journal.pone.0092205
- Gallagher N, Cardwell C, Hughes C, O'Reilly D. Increase in the pharmacological management of type 2 diabetes with pay-for-performance in primary care in the UK. *Diabet Med*. 2015;32:62-8. [PMID: 25185888] doi:10.1111/dme.12575
- Kendrick T, Stuart B, Newell C, Geraghty AW, Moore M. Did NICE guidelines and the Quality Outcomes Framework change GP antidepressant prescribing in England? Observational study with time trend analyses 2003-2013. *J Affect Disord*. 2015;186:171-7. [PMID: 26241666] doi:10.1016/j.jad.2015.06.052
- Kontopantelis E, Reeves D, Valderas JM, Campbell S, Doran T. Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study. *BMJ Qual Saf*. 2013;22:53-64. [PMID: 22918988] doi:10.1136/bmjqs-2012-001033
- MacBride-Stewart SP, Elton R, Walley T. Do quality incentives change prescribing patterns in primary care? An observational study in Scotland. *Fam Pract*. 2008;25:27-32. [PMID: 18245796] doi:10.1093/fampra/cmm074
- McLintock K, Russell AM, Alderson SL, West R, House A, Westerman K, et al. The effects of financial incentives for case finding for depression in patients with diabetes and coronary heart disease: interrupted time series analysis. *BMJ Open*. 2014;4:e005178. [PMID: 25142262] doi:10.1136/bmjopen-2014-005178
- Vamos EP, Pape UJ, Bottle A, Hamilton FL, Curcin V, Ng A, et al. Association of practice size and pay-for-performance incentives with

- the quality of diabetes management in primary care. *CMAJ*. 2011; 183:E809-16. [PMID: 21810950] doi:10.1503/cmaj.101187
33. Doran T, Kontopantelis E, Valderas JM, Campbell S, Roland M, Salisbury C, et al. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. *BMJ*. 2011;342:d3590. [PMID: 21712336] doi:10.1136/bmj.d3590
34. Harrison MJ, Dusheiko M, Sutton M, Gravelle H, Doran T, Roland M. Effect of a national primary care pay for performance scheme on emergency hospital admissions for ambulatory care sensitive conditions: controlled longitudinal study. *BMJ*. 2014;349:g6423. [PMID: 25389120] doi:10.1136/bmj.g6423
35. Karunaratne K, Stevens P, Irving J, Hobbs H, Kilbride H, Kingston R, et al. The impact of pay for performance on the control of blood pressure in people with chronic kidney disease stage 3-5. *Nephrol Dial Transplant*. 2013;28:2107-16. [PMID: 23658247] doi:10.1093/ndt/gft093
36. Sutton M, Elder R, Guthrie B, Watt G. Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers. *Health Econ*. 2010;19:1-13. [PMID: 19206084] doi:10.1002/hec.1440
37. Calvert M, Shankar A, McManus RJ, Lester H, Freemantle N. Effect of the Quality and Outcomes Framework on diabetes care in the United Kingdom: retrospective cohort study. *BMJ*. 2009;338:b1870. [PMID: 19474024] doi:10.1136/bmj.b1870
38. Hamilton FL, Lavery AA, Huckvale K, Car J, Majeed A, Millett C. Financial incentives and inequalities in smoking cessation interventions in primary care: before-and-after study. *Nicotine Tob Res*. 2016;18:341-50. [PMID: 25995158] doi:10.1093/ntr/ntv107
39. Kontopantelis E, Springate DA, Ashworth M, Webb RT, Buchan IE, Doran T. Investigating the relationship between quality of primary care and premature mortality in England: a spatial whole-population study. *BMJ*. 2015;350:h904. [PMID: 25733592] doi:10.1136/bmj.h904
40. Millett C, Bottle A, Ng A, Curcin V, Molokhia M, Saxena S, et al. Pay for performance and the quality of diabetes management in individuals with and without co-morbid medical conditions. *J R Soc Med*. 2009;102:369-77. [PMID: 19734534] doi:10.1258/jrsm.2009.090171
41. Murray J, Saxena S, Millett C, Curcin V, de Lusignan S, Majeed A. Reductions in risk factors for secondary prevention of coronary heart disease by ethnic group in south-west London: 10-year longitudinal study (1998-2007). *Fam Pract*. 2010;27:430-8. [PMID: 20538744] doi:10.1093/fampra/cm030
42. Norbury M, Fawkes N, Guthrie B. Impact of the GP contract on inequalities associated with influenza immunisation: a retrospective population-database analysis. *Br J Gen Pract*. 2011;61:e379-85. [PMID: 21722444] doi:10.3399/bjgp11X583146
43. Simpson CR, Hannaford PC, Ritchie LD, Sheikh A, Williams D. Impact of the pay-for-performance contract and the management of hypertension in Scottish primary care: a 6-year population-based repeated cross-sectional study. *Br J Gen Pract*. 2011;61:e443-51. [PMID: 21722469] doi:10.3399/bjgp11X583407
44. Smith CJ, Gribbin J, Challen KB, Hubbard RB. The impact of the 2004 NICE guideline and 2003 General Medical Services contract on COPD in primary care in the UK. *QJM*. 2008;101:145-53. [PMID: 18180254] doi:10.1093/qjmed/hcm155
45. Szatkowski L, McNeill A, Lewis S, Coleman T. A comparison of patient recall of smoking cessation advice with advice recorded in electronic medical records. *BMC Public Health*. 2011;11:291. [PMID: 21569283] doi:10.1186/1471-2458-11-291
46. Taggar JS, Coleman T, Lewis S, Szatkowski L. The impact of the Quality and Outcomes Framework (QOF) on the recording of smoking targets in primary care medical records: cross-sectional analyses from The Health Improvement Network (THIN) database. *BMC Public Health*. 2012;12:329. [PMID: 22559290] doi:10.1186/1471-2458-12-329
47. Tahrani AA, McCarthy M, Godson J, Taylor S, Slater H, Capps N, et al. Diabetes care and the new GMS contract: the evidence for a whole county. *Br J Gen Pract*. 2007;57:483-5. [PMID: 17550674]
48. Vaghela P, Ashworth M, Schofield P, Gulliford MC. Population intermediate outcomes of diabetes under pay-for-performance incentives in England from 2004 to 2008. *Diabetes Care*. 2009;32:427-9. [PMID: 19106379] doi:10.2337/dc08-1999
49. Mason T, Sutton M, Whittaker W, McSweeney T, Millar T, Donmall M, et al. The impact of paying treatment providers for outcomes: difference-in-differences analysis of the 'payment by results for drugs recovery' pilot. *Addiction*. 2015;110:1120-8. [PMID: 26058447] doi:10.1111/add.12920
50. Kalwij S, French S, Mugezi R, Baraitser P. Using educational outreach and a financial incentive to increase general practices' contribution to chlamydia screening in South-East London 2003-2011. *BMC Public Health*. 2012;12:802. [PMID: 22984897] doi:10.1186/1471-2458-12-802
51. Chang RE, Lin SP, Aron DC. A pay-for-performance program in Taiwan improved care for some diabetes patients, but doctors may have excluded sicker ones. *Health Aff (Millwood)*. 2012;31:93-102. [PMID: 22232099] doi:10.1377/hlthaff.2010.0402
52. Chen YC, Lee CT, Lin BJ, Chang YY, Shi HY. Impact of pay-for-performance on mortality in diabetes patients in Taiwan: a population-based study. *Medicine (Baltimore)*. 2016;95:e4197. [PMID: 27399144] doi:10.1097/MD.00000000000004197
53. Chen CC, Cheng SH. Does pay-for-performance benefit patients with multiple chronic conditions? Evidence from a universal coverage health care system. *Health Policy Plan*. 2016;31:83-90. [PMID: 25944704] doi:10.1093/heapol/czv024
54. Cheng SH, Lee TT, Chen CC. A longitudinal examination of a pay-for-performance program for diabetes care: evidence from a natural experiment. *Med Care*. 2012;50:109-16. [PMID: 22249920] doi:10.1097/MLR.0b013e31822d5d36
55. Hsieh HM, Tsai SL, Shin SJ, Mau LW, Chiu HC. Cost-effectiveness of diabetes pay-for-performance incentive designs. *Med Care*. 2015; 53:106-15. [PMID: 25397966] doi:10.1097/MLR.0000000000000264
56. Hsieh HM, Lin TH, Lee IC, Huang CJ, Shin SJ, Chiu HC. The association between participation in a pay-for-performance program and macrovascular complications in patients with type 2 diabetes in Taiwan: a nationwide population-based cohort study. *Prev Med*. 2016;85:53-9. [PMID: 26740347] doi:10.1016/j.ypmed.2015.12.013
57. Liao PJ, Lin TY, Wang TC, Ting MK, Wu IW, Huang HT, et al. Long-term and interactive effects of pay-for-performance interventions among diabetic nephropathy patients at the early chronic kidney disease stage. *Medicine (Baltimore)*. 2016;95:e3282. [PMID: 27057892] doi:10.1097/MD.00000000000003282
58. Lee TT, Cheng SH, Chen CC, Lai MS. A pay-for-performance program for diabetes care in Taiwan: a preliminary assessment. *Am J Manag Care*. 2010;16:65-9. [PMID: 20148607]
59. Tan EC, Pwu RF, Chen DR, Yang MC. Is a diabetes pay-for-performance program cost-effective under the National Health Insurance in Taiwan? *Qual Life Res*. 2014;23:687-96. [PMID: 23975377] doi:10.1007/s11136-013-0502-x
60. Lee CY, Chi MJ, Yang SL, Lo HY, Cheng SH. Using financial incentives to improve the care of tuberculosis patients. *Am J Manag Care*. 2015;21:e35-42. [PMID: 25880266]
61. Li YH, Tsai WC, Khan M, Yang WT, Lee TF, Wu YC, et al. The effects of pay-for-performance on tuberculosis treatment in Taiwan. *Health Policy Plan*. 2010;25:334-41. [PMID: 20207703] doi:10.1093/heapol/czq006
62. Chen HJ, Huang N, Chen LS, Chou YJ, Li CP, Wu CY, et al. Does pay-for-performance program increase providers adherence to guidelines for managing hepatitis B and hepatitis C virus infection in Taiwan? *PLoS One*. 2016;11:e0161002. [PMID: 27517172] doi:10.1371/journal.pone.0161002
63. Martens JD, Werkhoven MJ, Severens JL, Winkens RA. Effects of a behaviour independent financial incentive on prescribing behaviour of general practitioners. *J Eval Clin Pract*. 2007;13:369-73. [PMID: 17518801]
64. Pechlivanoglou P, Wieringa JE, de Jager T, Postma MJ. The effect of financial and educational incentives on rational prescribing. A state-space approach. *Health Econ*. 2015;24:439-53. [PMID: 24519732] doi:10.1002/hec.3030

65. Kirschner K, Braspenning J, Akkermans RP, Jacobs JE, Grol R. Assessment of a pay-for-performance program in primary care designed by target users. *Fam Pract*. 2013;30:161-71. [PMID: 22997223] doi:10.1093/fampra/cms055
66. Kiran T, Wilton AS, Moineddin R, Paszat L, Glazier RH. Effect of payment incentives on cancer screening in Ontario primary care. *Ann Fam Med*. 2014;12:317-23. [PMID: 25024239] doi:10.1370/afm.1664
67. Li J, Hurley J, DeCicca P, Buckley G. Physician response to pay-for-performance: evidence from a natural experiment. *Health Econ*. 2014;23:962-78. [PMID: 23861240] doi:10.1002/hec.2971
68. Michel-Lepage A, Ventelou B. The true impact of the French pay-for-performance program on physicians' benzodiazepines prescription behavior. *Eur J Health Econ*. 2016;17:723-32. [PMID: 26304210] doi:10.1007/s10198-015-0717-6
69. Sicsic J, Franc C. Impact assessment of a pay-for-performance program on breast cancer screening in France using micro data. *Eur J Health Econ*. 2016. [PMID: 27329654]
70. Rat C, Penhouet G, Gaultier A, Chaslerie A, Pivette J, Nguyen JM, et al. Did the new French pay-for-performance system modify benzodiazepine prescribing practices? *BMC Health Serv Res*. 2014;14:301. [PMID: 25011548] doi:10.1186/1472-6963-14-301
71. Greene J. An examination of pay-for-performance in general practice in Australia. *Health Serv Res*. 2013;48:1415-32. [PMID: 23350933] doi:10.1111/1475-6773.12033
72. Zuckerman RB, Sheingold SH, Orav EJ, Ruhter J, Epstein AM. Readmissions, observation, and the Hospital Readmissions Reduction Program. *N Engl J Med*. 2016;374:1543-51. [PMID: 26910198] doi:10.1056/NEJMsa1513024
73. Figueroa JF, Tsugawa Y, Zheng J, Orav EJ, Jha AK. Association between the Value-Based Purchasing pay for performance program and patient mortality in US hospitals: observational study. *BMJ*. 2016;353:i2214. [PMID: 27160187] doi:10.1136/bmj.i2214
74. Ryan AM, Burgess JF Jr, Pesko MF, Borden WB, Dimick JB. The early effects of Medicare's mandatory hospital pay-for-performance program. *Health Serv Res*. 2015;50:81-97. [PMID: 25040485] doi:10.1111/1475-6773.12206
75. Ryan A, Sutton M, Doran T. Does winning a pay-for-performance bonus improve subsequent quality performance? Evidence from the Hospital Quality Incentive Demonstration. *Health Serv Res*. 2014;49:568-87. [PMID: 23909992] doi:10.1111/1475-6773.12097
76. Andriole KP, Prevedello LM, Dufault A, Pezeshk P, Bransfield R, Hanson R, et al. Augmenting the impact of technology adoption with financial incentive to improve radiology report signature times. *J Am Coll Radiol*. 2010;7:198-204. [PMID: 20193925] doi:10.1016/j.jacr.2009.11.011
77. Benzer JK, Young GJ, Burgess JF Jr, Baker E, Mohr DC, Charns MP, et al. Sustainability of quality improvement following removal of pay-for-performance incentives. *J Gen Intern Med*. 2014;29:127-32. [PMID: 23929219] doi:10.1007/s11606-013-2572-4
78. Kristensen SR, Meacock R, Turner AJ, Boaden R, McDonald R, Roland M, et al. Long-term effect of hospital pay for performance on mortality in England. *N Engl J Med*. 2014;371:540-8. [PMID: 25099578] doi:10.1056/NEJMoa1400962
79. Kuo RN, Chung KP, Lai MS. Effect of the Pay-for-Performance Program for Breast Cancer Care in Taiwan. *J Oncol Pract*. 2011;7:e8s-e15s. [PMID: 21886513] doi:10.1200/JOP.2011.000314
80. Vermeulen MJ, Stukel TA, Boozary AS, Guttman A, Schull MJ. The effect of pay for performance in the emergency department on patient waiting times and quality of care in Ontario, Canada: a difference-in-differences analysis. *Ann Emerg Med*. 2016;67:496-505. [PMID: 26215670] doi:10.1016/j.annemergmed.2015.06.028
81. Colais P, Pinnarelli L, Fusco D, Davoli M, Braga M, Perucci CA. The impact of a pay-for-performance system on timing to hip fracture surgery: experience from the Lazio Region (Italy). *BMC Health Serv Res*. 2013;13:393. [PMID: 24099264] doi:10.1186/1472-6963-13-393
82. Yang JH, Kim SM, Han SJ, Knaak M, Yang GH, Lee KD, et al. The impact of Value Incentive Program (VIP) on the quality of hospital care for acute stroke in Korea. *Int J Qual Health Care*. 2016;28:580-585. [PMID: 27650012]
83. Gillam S, Steel N. The Quality and Outcomes Framework—where next? *BMJ*. 2013;346:f659. [PMID: 23393112] doi:10.1136/bmj.f659
84. Tricco AC, Ivers NM, Grimshaw JM, Moher D, Turner L, Galipeau J, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. *Lancet*. 2012;379:2252-61. [PMID: 22683130] doi:10.1016/S0140-6736(12)60480-2
85. Emanuel EJ, Ubel PA, Kessler JB, Meyer G, Muller RW, Navathe AS, et al. Using behavioral economics to design physician incentives that deliver high-value care. *Ann Intern Med*. 2016;164:114-9. [PMID: 26595370] doi:10.7326/M15-1330
86. Kondo KK, Damberg CL, Mendelson A, Motúapuaka M, Freeman M, O'Neil M, et al. Implementation processes and pay for performance in healthcare: a systematic review. *J Gen Intern Med*. 2016;31 Suppl 1:61-9. [PMID: 26951276] doi:10.1007/s11606-015-3567-0
87. Kondo K, Damberg C, Mendelson A, Motúapuaka M, Freeman M, O'Neil M, et al. Understanding the Intervention and Implementation Factors Associated with Benefits and Harms of Pay for Performance Programs in Healthcare. Washington, DC: U.S. Department of Veterans Affairs; 2015. [PMID: 27054229]
88. Kansagara D, Tuepker A, Joos S, Nicolaidis C, Skaperdas E, Hickam D. Getting performance metrics right: a qualitative study of staff experiences implementing and measuring practice transformation. *J Gen Intern Med*. 2014;29 Suppl 2:S607-13. [PMID: 24557515] doi:10.1007/s11606-013-2764-y
89. Powell AA, White KM, Partin MR, Halek K, Christianson JB, Neil B, et al. Unintended consequences of implementing a national performance measurement system into local practice. *J Gen Intern Med*. 2012;27:405-12. [PMID: 21993998] doi:10.1007/s11606-011-1906-3
90. Kizer KW, Kirsh SR. The double edged sword of performance measurement [Editorial]. *J Gen Intern Med*. 2012;27:395-7. [PMID: 22271270] doi:10.1007/s11606-011-1981-5
91. Casalino LP, Gans D, Weber R, Cea M, Tuchovsky A, Bishop TF, et al. US physician practices spend more than \$15.4 billion annually to report quality measures. *Health Aff (Millwood)*. 2016;35:401-6. [PMID: 26953292] doi:10.1377/hlthaff.2015.1258
92. Roland M, Campbell S. Successes and failures of pay for performance in the United Kingdom. *N Engl J Med*. 2014;370:1944-9. [PMID: 24827040] doi:10.1056/NEJMhpr1316051
93. Parast L, Doyle B, Damberg CL, Shetty K, Ganz DA, Wenger NS, et al. Challenges in assessing the process-outcome link in practice. *J Gen Intern Med*. 2015;30:359-64. [PMID: 25564435] doi:10.1007/s11606-014-3150-0
94. Findlay S, Berenson R, Lott R, Gnadinger T. Health Policy Brief: Medicare's New Physician Payment System. A 2015 law has the potential to transform how Medicare pays physicians. *Health Affairs*. 21 April 2016.

Current Author Addresses: Mr. Mendelson: Oregon Health & Science University, Mail Code MDYCHSE, 3181 SW Sam Jackson Park Road, Portland, OR 97239.

Drs. Kondo, O'Neil, and Kansagara; Ms. Low; Ms. Motúapuaka; Ms. Freeman; and Ms. Relevo: VA Portland Health Care System, Mail Code R&D 71, 3710 SW US Veterans Hospital Road, Portland, OR 97239-2999.

Dr. Damberg: RAND (Health), 1776 Main Street, Santa Monica, CA 90407-2138.

Author Contributions: Conception and design: K. Kondo, M. Motúapuaka, M. O'Neil, R. Relevo, D. Kansagara.

Analysis and interpretation of the data: A. Mendelson, K. Kondo, C. Damberg, A. Low, M. Motúapuaka, M. Freeman, M. O'Neil, D. Kansagara.

Drafting of the article: A. Mendelson, K. Kondo, C. Damberg, M. Motúapuaka, M. O'Neil, D. Kansagara.

Critical revision of the article for important intellectual content: A. Mendelson, K. Kondo, A. Low, M. Motúapuaka, D. Kansagara.

Final approval of the article: A. Mendelson, K. Kondo, C. Damberg, A. Low, M. Motúapuaka, M. Freeman, M. O'Neil, R. Relevo, D. Kansagara.

Obtaining of funding: D. Kansagara.

Administrative, technical, or logistic support: A. Low, M. Motúapuaka, M. Freeman.

Collection and assembly of data: A. Mendelson, K. Kondo, C. Damberg, A. Low, M. Motúapuaka, M. Freeman, M. O'Neil, R. Relevo, D. Kansagara.