Cochrane
Library

Cochrane Database of Systematic Reviews

Pay for performance for hospitals (Review)

Mathes T, Pieper D, Morche J, Polus S, Jaschinski T, Eikermann M

Mathes T, Pieper D, Morche J, Polus S, Jaschinski T, Eikermann M.
Pay for performance for hospitals.
Cochrane Database of Systematic Reviews 2019, Issue 7. Art. No.: CD011156.
DOI: [10.1002/14651858.CD011156.pub2](https://doi.org/10.1002/14651858.CD011156.pub2).

www.cochranelibrary.com

Pay for performance for hospitals (Review)

Copyright © 2019 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

WILEY

TABLE OF CONTENTS

HEADER	1
ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
SUMMARY OF FINDINGS	4
BACKGROUND	9
OBJECTIVES	10
METHODS	10
RESULTS	13
Figure 1.	14
Figure 2.	16
DISCUSSION	19
AUTHORS' CONCLUSIONS	20
ACKNOWLEDGEMENTS	20
REFERENCES	21
CHARACTERISTICS OF STUDIES	25
ADDITIONAL TABLES	84
APPENDICES	97
CONTRIBUTIONS OF AUTHORS	105
DECLARATIONS OF INTEREST	105
SOURCES OF SUPPORT	105
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	105
INDEX TERMS	106

[Intervention Review]

Pay for performance for hospitals

Tim Mathes¹, Dawid Pieper¹, Johannes Morche², Stephanie Polus¹, Thomas Jaschinski¹, Michaela Eikermann³

¹Institute for Research in Operative Medicine (IFOM) - Department for Evidence-based Health Services Research, University Witten/Herdecke, Cologne, Germany. ²Medical Consultancy Department, Federal Joint Committee, Berlin, Germany. ³Department of Evidence-based medicine, Medical advisory service of social health insurance (MDS), Essen, Germany

Contact address: Tim Mathes, Institute for Research in Operative Medicine (IFOM) - Department for Evidence-based Health Services Research, University Witten/Herdecke, Ostmerheimer Str. 200 (House 38), Cologne, 51109, Germany. tim.mathes@uni-wh.de.

Editorial group: Cochrane Effective Practice and Organisation of Care Group.

Publication status and date: New, published in Issue 7, 2019.

Citation: Mathes T, Pieper D, Morche J, Polus S, Jaschinski T, Eikermann M. Pay for performance for hospitals. *Cochrane Database of Systematic Reviews* 2019, Issue 7. Art. No.: CD011156. DOI: [10.1002/14651858.CD011156.pub2](https://doi.org/10.1002/14651858.CD011156.pub2).

Copyright © 2019 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Pay-for-Performance (P4P) is a payment model that rewards health care providers for meeting pre-defined targets for quality indicators or efficacy parameters to increase the quality or efficacy of care.

Objectives

Our objective was to assess the impact of P4P for in-hospital delivered health care on the quality of care, resource use and equity. Our objective was not only to answer the question whether P4P works in general (simple perspective) but to provide a comprehensive and detailed overview of P4P with a focus on analyzing the intervention components, the context factors and their interrelation (more complex perspective).

Search methods

We searched CENTRAL, MEDLINE, Embase, three other databases and two trial registers on 27 June 2018. In addition, we searched conference proceedings, gray literature and web pages of relevant health care institutions, contacted experts in the field, conducted cited reference searches and performed cross-checks of included references and systematic reviews on the same topic.

Selection criteria

We included randomized trials, cluster randomized trials, non-randomized clustered trials, controlled before-after studies, interrupted time series and repeated measures studies that analyzed hospitals, hospital units or groups of hospitals and that compared any kind of P4P to a basic payment scheme (e.g. capitation) without P4P. Studies had to analyze at least one of the following outcomes to be eligible: patient outcomes; quality of care; utilization, coverage or access; resource use, costs and cost shifting; healthcare provider outcomes; equity; adverse effects or harms.

Data collection and analysis

Two review authors independently screened all citations for inclusion, extracted study data and assessed risk of bias for each included study. Study characteristics were extracted by one reviewer and verified by a second.

We did not perform meta-analysis because the included studies were too heterogenous regarding hospital characteristics, the design of the P4P programs and study design. Instead we present a structured narrative synthesis considering the complexity as well as the context/setting of the intervention. We assessed the certainty of evidence using the GRADE approach and present the results narratively in 'Summary of findings' tables.

Main results

We included 27 studies (20 CBA, 7 ITS) on six different P4P programs. Studies analyzed between 10 and 4267 centers. All P4P programs targeted acute or emergency physical conditions and compared a capitation-based payment scheme without P4P to the same capitation-based payment scheme combined with a P4P add-on. Two P4P programs used rewards or penalties; one used first rewards and then penalties; two used penalties only and one used rewards only. Four P4P programs were established and evaluated in the USA, one in England and one in France.

Most studies showed no difference or a very small effect in favor of the P4P program. The impact of each P4P program was as follows.

Premier Hospital Quality Incentive Demonstration Program: It is uncertain whether this program, which used rewards for some hospitals and penalties for others, has an impact on mortality, adverse clinical events, quality of care, equity or resource use as the certainty of the evidence was very low.

Value-Based Purchasing Program: It is uncertain whether this program, which used rewards for some hospitals and penalties for others, has an impact on mortality, adverse clinical events or quality of care as the certainty of the evidence was very low. Equity and resource use outcomes were not reported in the studies, which evaluated this program.

Non-payment for Hospital-Acquired Conditions Program: It is uncertain whether this penalty-based program has an impact on adverse clinical events as the certainty of the evidence was very low. Mortality, quality of care, equity and resource use outcomes were not reported in the studies, which evaluated this program.

Hospital Readmissions Reduction Program: None of the studies that examined this penalty-based program reported mortality, adverse clinical events, quality of care (process quality score), equity or resource use outcomes.

Advancing Quality Program: It is uncertain whether this reward-/penalty-based program has an impact on mortality as the certainty of the evidence was very low. Adverse clinical events, quality of care, equity and resource use outcomes were not reported in any study.

Financial Incentive to Quality Improvement Program: It is uncertain whether this reward-based program has an impact on quality of care, as the certainty of the evidence was very low. Mortality, adverse clinical events, equity and resource use outcomes were not reported in any study.

Subgroup analysis (analysis of modifying design and context factors)

Analysis of P4P design factors provides some hints that non-payments compared to additional payments and payments for quality attainment (e.g. falling below specified mortality threshold) compared to quality improvement (e.g. reduction of mortality by specified percent points within one year) may have a stronger impact on performance.

Authors' conclusions

It is uncertain whether P4P, compared to capitation-based payments without P4P for hospitals, has an impact on patient outcomes, quality of care, equity or resource use as the certainty of the evidence was very low (or we found no studies on the outcome) for all P4P programs. The effects on patient outcomes of P4P in hospitals were at most small, regardless of design factors and context/setting. It seems that with additional payments only small short-term but non-sustainable effects can be achieved. Non-payments seem to be slightly more effective than bonuses and payments for quality attainment seem to be slightly more effective than payments for quality improvement.

PLAIN LANGUAGE SUMMARY

Pay for performance (payment or penalty methods to encourage hospitals to increase quality of care)

What was the aim of this review

The aim of this Cochrane Review was to find out if 'pay for performance' — that is, providing hospitals with monetary incentives to meet targets or penalizing them for failing to reach those targets — can improve the quality of patient care, resource use and equity. The review was not limited to a certain health problem. Cochrane researchers collected and analyzed all relevant studies to answer this question.

Key messages

Pay for performance improved patient outcomes (mortality, clinical adverse events) either only very slightly or not at all. It seems that providing hospitals with additional payments to reward performance achieves only small, short-term, but non-sustainable, effects. Penalizing hospitals through non-payment for failure to reach performance targets seems to be slightly more effective than providing additional payments for performance; and payments for quality attainment (e.g. falling below specified mortality threshold) seem to be slightly more effective than payments for quality improvement (e.g. reduction of mortality by specified percent points within one year). It was not possible to determine if pay for performance affects patient outcomes because the certainty of the available evidence was judged to be very low. The impact of pay for performance on equity is unclear.

What was studied in the review?

The payment method for reimbursing health care delivered in hospitals can have an impact on patient outcomes, quality of care, equity, utilization, health care provider outcomes (e.g. workload) and adverse effects. The Cochrane researchers assessed the impact of payment methods that are based on hospital performance (e.g. hospital-acquired infections) and are thereby aiming to stimulate an improvement of hospital performance.

Main results

The review authors found 27 relevant studies that compared six different P4P programs. Twenty-four were from the USA, two from the UK and one was from France. All studies compared pay for performance (through the use of rewards or penalties or both) with no pay for performance, i.e. a basic payment scheme without a component that incentivizes quality of care. The studies were either funded by government agencies or received no funding.

There was no improvement of patient outcomes (mortality, adverse clinical events) or the improvement was at most very small. Consequently, we are uncertain whether P4P has a positive impact on patient outcomes because the certainty of the evidence was very low. There was a slightly larger improvement in quality of care. Non-payments ('sticks') seem to be a little bit more effective than additional payments ('carrots'). The impact of Pay-for-Performance on equity is unclear. We found no data on utilization (resource use), health care provider outcomes (quality of care) and adverse effects.

All studies were performed in high-income countries (USA, UK, France). Because of differences in the health care systems and the complexity of the Pay-for-Performance programs, the applicability of our findings to other countries is limited.

Future studies should put a stronger focus on the features that might modify the effect of P4P. In particular, the interaction of P4P features and context/setting (e.g. larger incentives for hospitals in a bad financial situation) should be evaluated.

How up to date is this review?

We searched for studies that had been published up to 27 June 2018.

SUMMARY OF FINDINGS

Summary of findings for the main comparison. Premier Hospital Quality Incentive Demonstration compared to capitation without P4P for hospitals

Premier Hospital Quality Incentive Demonstration compared to Capitation without P4P for Hospitals

Patient or population: Hospitals

Setting: USA

Intervention: Premier Hospital Quality Incentive Demonstration: Rewards or penalties for quality attainment and improvement depending on process quality score compared to the national average; voluntary participation

Comparison: Capitation without P4P

Outcomes	Impact	Nº of hospitals (studies)	Certainty of the evidence (GRADE)
Mortality follow up: 30 days	Studies showed no or a very small reduction in overall mortality in the hospitals under Premier Hospital Quality Incentive Demonstration. Condition-specific mortality was reduced for most conditions but effects varied for different conditions and was heterogeneous across studies.	5126 (3 CBA)	⊕⊕⊕⊕ ^{a, b} VERY LOW
Adverse clinical events follow up: inpatient	Premier Hospital Quality Incentive Demonstration reduced adverse clinical events slightly (OR for reduction 1.11, 95% CI 0.91 to 1.36)	1139 (1 CBA)	⊕⊕⊕⊕ ^{a, b} VERY LOW
Quality of care (process quality score) follow up: not applicable	Premier Hospital Quality Incentive Demonstration improved quality of care slightly	4249 (3 CBA)	⊕⊕⊕⊕ ^{a, b} VERY LOW
Equity (differences between ethnic groups in access to recommended care) follow up: not applicable	Premier Hospital Quality Incentive Demonstration increased differences between ethnic groups slightly (access to recommended care %: white vs non-white: -0.6)	1063 (1 CBA)	⊕⊕⊕⊕ ^{a, b} VERY LOW
Resource use (mean costs per admission) follow up: not applicable	Premier Hospital Quality Incentive Demonstration had no impact on hospital costs.	1392 (2 CBA)	⊕⊕⊕⊕ ^{a, b} VERY LOW

^aDowngraded 1 level for risk of bias

^bDowngraded 1 level for imprecision

GRADE Working Group grades of evidence

High certainty: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Summary of findings 2. Value-Based Purchasing compared to capitation without P4P for hospitals

Value-Based Purchasing compared to Capitation without P4P for Hospitals

Patient or population: Hospitals

Setting: USA

Intervention: Value-Based Purchasing: Rewards or penalties for quality attainment and improvement depending on clinical processes, patient experience, patient outcomes and resources compared to the national average and baseline quality; obligatory participation

Comparison: Capitation without P4P

Outcomes	Impact	Nº of hospitals (studies)	Certainty of the evidence (GRADE)
Mortality follow up: 30 days	Small reduction of mortality in hospitals under Values-Based Purchasing	6248 (2 CBA)	⊕⊕⊕⊕ ^{a, b} VERY LOW
Adverse clinical events	No studies on this outcome identified		
Quality of care (process quality score) follow up: not applicable	Values-Based Purchasing improved quality of processes slightly	5358 (2 CBA)	⊕⊕⊕⊕ ^{a, b} VERY LOW
Equity	No studies on this outcome identified		
Resource use	No studies on this outcome identified		

^aDowngraded 1 level for risk of bias

^bDowngraded 1 level for imprecision

GRADE Working Group grades of evidence

High certainty: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Summary of findings 3. Non-payment for Hospital-Acquired Conditions Program compared to capitation without P4P for hospitals

Non-payment for Hospital-Acquired Conditions Program compared to Capitation without P4P for Hospitals

Patient or population: Hospitals

Setting: USA

Intervention: Non-payment for Hospital-Acquired Conditions Program; No payment for conditions acquired in hospital (penalties); obligatory participation

Comparison: Capitation without P4P

Outcomes	Impact	Nº of hospitals (studies)	Certainty of the evidence (GRADE)
Mortality	No studies on this outcome identified		

Adverse clinical events* follow up: in hospital	Non-payment for hospital-acquired conditions reduced hospital acquired conditions	> 3568# (5 ITS, 1 CBA)	⊕⊕⊕⊕ ^a VERY LOW
Quality of care	No studies on this outcome identified		
Equity	No studies on this outcome identified		
Resource use	No studies on this outcome identified		

*Surgical side infections; central catheter-associated bloodstream infections; catheter-associated urinary tract infections; hospital acquired pressure ulcers; falls

#in 1 study no information on the number of included hospitals was given

^aDowngraded for risk of bias

GRADE Working Group grades of evidence

High certainty: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Summary of findings 4. Hospital Readmissions Reduction Program compared to capitation without P4P for hospitals

Hospital Readmissions Reduction Program compared to Capitation without P4P for Hospitals

Patient or population: Hospitals

Setting: USA

Intervention: Hospital Readmissions Reduction Program: No payments for readmissions for the same reason within 30 days above the national average (penalties); obligatory participation

Comparison: Capitation without P4P

Outcomes	Impact	Nº of hospitals (studies)	Certainty of the evidence (GRADE)
Mortality	No studies on this outcome identified		
Adverse clinical events	No studies on this outcome identified		
Quality of care	No studies on this outcome identified		
Equity	No studies on this outcome identified		
Resource use	No studies on this outcome identified		

^aDowngraded 1 level for risk of bias

^bDowngraded 1 level for imprecision

GRADE Working Group grades of evidence

High certainty: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Summary of findings 5. Advancing Quality Program compared to capitation without P4P for hospitals

Advancing Quality Program compared to Capitation without P4P for Hospitals

Patient or population: Hospitals

Setting: England

Intervention: Advancing Quality Program: 1. Phase: bonuses if a certain quality threshold (clinical process outcomes, patient outcomes, patient experience) was reached (quality improvement); Phase 2: payments withhold until a certain quality threshold (clinical process outcomes, patient outcomes, patient experience) was reached (quality attainment); voluntary participation

Comparison: Capitation without P4P

Outcomes	Impact	Nº of hospitals (studies)	Certainty of the evidence (GRADE)
Mortality follow up: 30 days	Short term reduction in mortality that could not be sustained in the long run	156 (2 CBA)	⊕⊕⊕⊕ ^{a, b} VERY LOW
Adverse clinical events	No studies on this outcome identified		
Quality of care	No studies on this outcome identified		
Equity	No studies on this outcome identified		
Resource use	No studies on this outcome identified		

^aDowngraded 1 level for risk of bias

^bDowngraded 1 level for imprecision

GRADE Working Group grades of evidence

High certainty: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Summary of findings 6. Financial Incentive to Quality Improvement compared to capitation without P4P for hospitals

Financial Incentive to Quality Improvement compared to Capitation without P4P for Hospitals

Patient or population: Hospitals

Setting: France

Intervention: Financial Incentive to Quality Improvement: Rewards for quality attainment and improvement depending on quality targets (process quality score) compared to the national median; voluntary participation

Comparison: Capitation without P4P

Outcomes	Impact	Nº of hospitals (studies)	Certainty of the evidence (GRADE)
Mortality	No studies on this outcome identified		
Adverse clinical events	No studies on this outcome identified		
Quality of care (process quality score) follow up: not applicable	Financial Incentive to Quality Improvement improved quality of processes slightly	377 (1 CBA)	⊕⊕⊕⊕ ^{a, b} VERY LOW
Equity	No studies on this outcome identified		
Resource use	No studies on this outcome identified		

^aDowngraded 1 level for risk of bias

^bDowngraded 1 level for imprecision

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

BACKGROUND

Hospital costs comprise a large share of the total healthcare budget. Payment methods for hospitals can influence organization behavior, which might affect quality and efficiency of provision of care. Providers can potentially be incentivized to deliver care that maximizes the patient's benefit while keeping resource use under control. Incentivizing providers to deliver high-quality care is especially important in healthcare markets because it is difficult for patients to judge quality of care and consequently they often have to rely on the decisions of providers (information asymmetries) (Blomqvist 1991). In the past, policy interventions on reimbursing methods for hospital-delivered healthcare have been implemented with the primary aim of cost containment in many countries (Abel-Smith 1994; Böcking 2005; Dixon 2004; Draper 2006). More recently, some countries have implemented payment methods with the primary aim to increase performance (quality, efficiency, or both) (Kondo 2016).

Description of the intervention

There are six main types of reimbursement schemes (WHO 2013): fee for service (pay per procedure); pay per diem; case-based reimbursement; capitation; line item budget; and global budget. P4P are rewards for meeting pre-defined targets for quality indicators or efficacy parameters. P4P is not a payment method itself but an add-on to these reimbursement schemes, which is meant to create an incentive for increasing the quality of care (Lindenauer 2007). The payments are directly linked to quality targets (e.g. reduction in mortality, care according clinical practice guidelines, reduction in avoidable complications). P4P can have different designs (Eijkenaar 2013); and these designs can differ in the following features.

- Payment at group level (e.g. hospital, units) or individual level (e.g. physicians).
- Rewards (additional payments) or penalties (non-payments or repayments).
- Size of payments.
- Payments for quality attainment (e.g. falling below specified mortality threshold) or for quality improvement (e.g. reduction of mortality by specified percent points within one year).
- Fixed payments (pre-defined amount for pre-defined targets) or relative payments (actual amount depends on the performance of other hospitals).
- Frequency of quality monitoring to assess performance.
- Frequency of payments.
- Payments linked to process or outcome quality.
- Obligatory or voluntary participation.
- Cover only a certain type of hospital (e.g. acute care) or ward (e.g. intensive care units) or all hospitals within a health care system.

Moreover, the different individual components might be mixed (e.g. incentives for some indications and penalties for other indications) or changed over time.

How the intervention might work

Payment methods should be designed to balance the advantages and disadvantages regarding conflicting objectives (costs and quality) and the interests of the different stakeholders (Zweifel 2009). However, each payment method also comes with potential

unintended effects. For example, in case-based reimbursement systems there is a risk that hospitals increase the number of admissions (WHO 2013), offer fewer services per case (leading to, for example, premature discharge) and are encouraged to practice case selection (patient/case mix optimizing), which may have an influence on access to care (e.g. 'cream skimming', that is providing care primarily for low-cost patients). For very ill patients especially, this can be problematic and may decrease quality of care (undersupply). Furthermore, there may be no incentive to increase the quality of services. In addition, patient demand for health care may be unresponsive to the quality of care because of the underlying information asymmetries (Petersen 2006).

For these reasons there might be room for quality improvement. The purpose of paying incentives for reaching target agreements is to raise quality, efficacy or both. P4P can improve health care quality in two ways. First, P4P might incentivize a volume/intensity increase or reduction of health care services which might result in a reduction of oversupply or undersupply. In the case of volume/intensity changes there is a direct effect on resource use. Second, P4P might incentivize the implementation of quality measures (e.g. more money for postgraduate training, more money for equipment or facilities). Theoretically, it is argued that P4P has the potential to raise the quality of care (Epstein 2012). However, there is a risk that hospitals aim to avoid treating sicker patients (risk selection) or patients with co-morbidities (cream skimming) to achieve better performance measures and thus receive additional payments instead of actually increasing the quality of care. Moreover, hospitals might focus only on the improvement of incentivized procedures/indications and neglect procedures/indications which are not incentivized.

Why it is important to do this review

Econometric models have shown an effect of payment methods on hospital volume/intensity (Ellis 1986; Ellis 1996; Ma 1994). In these econometric models, the utilization of services is typically the result of joint decisions between providers (e.g. hospitals) and patients (Jon 2012). In a "perfect" healthcare system, all hospitals would precisely provide the volume/intensity that perfectly balances quality of care and efficacy. In reality, however, this equilibrium is difficult to reach because involved stakeholders (payers, providers, patients) may pursue conflicting interests regarding access to care, profitability, expenses, cost containment, safety, quality, convenience, patient-centeredness and satisfaction (Porter 2010). Moreover, in reality there are many possibly influencing factors in addition to the payment method, for instance the influence of different organizational levels (e.g. units, hospital trusts) on hospital behavior. The response to payment methods can be further influenced by context and setting factors that might not be externally controllable or observable (e.g. professional ethics). Finally, the results of payment methods like volume/intensity, quality of care and costs are interrelated and the implementation of payment methods can be difficult (Craig 2008). Thus in practice the precise consequences of changes in hospital payment methods are difficult to predict — and even more so for quality of care and healthcare costs (Campbell 2007; Trochim 2006). In addition, the introduction of new payment methods can be accompanied by other interventions, such as hospital monitoring, which makes the prediction even more difficult. In view of this complexity, it is necessary to evaluate the real world impact of Pay-for-Performance (P4P) programs in hospitals to better inform future policy decisions.

Several systematic reviews have analyzed the effects of different payment methods. [Gosden 2000](#) found some hints that the payment of primary care physicians has an influence on the delivery of care. [Chaix-Couturier 2000](#) showed an association between the payment of primary care physicians with volume/intensity and various risks (e.g. limited access to care). [Witter 2012](#) found no clear effect of P4P in low- and middle-income countries; while Petersen and colleagues found small effects for P4P at the group level (e.g. hospitals) and concluded that the success of financial incentives requires careful design and that ongoing monitoring is critical to determine the effectiveness of financial incentives and their possible unintended effects on quality of care ([Petersen 2006](#)). Systematic reviews that try to answer whether P4P in hospitals is effective have mostly shown no or little effect ([Mehrotra 2009](#); [Mendelson 2017](#)).

OBJECTIVES

To assess the impact of P4P for in-hospital delivered health care on the quality of care, resource use and equity.

Our objective was not only to answer the question whether P4P works in general (simple perspective) but to provide a comprehensive and detailed overview of P4P with a focus on analyzing the intervention components, the context factors and their interrelation (more complex perspective) ([Petticrew 2013](#)).

METHODS

Criteria for considering studies for this review

Types of studies

We included studies in which a manipulation of, or change in, the payment method was analyzed and that applied the following study designs.

- Randomized (controlled) trials.
- Non-randomized (controlled) trials.
- Cluster-randomized trials with at least two intervention and control sites.
- Non-randomized cluster trials with at least two intervention and control sites.
- Controlled before-after (CBA) studies with at least two intervention and control sites.
- Interrupted time series (ITS) that have a clearly defined point in time when the intervention occurred and at least three data points before and three after the intervention.
- Repeated measure studies (RMS), which is an ITS study where measurements are made in the same individuals at each time point.

We included head-to-head comparisons of all possible pairs of interventions. We determined the study design using the Cochrane Effective Practice and Organisation of Care Group (EPOC) algorithm ([EPOC 2013a](#)). We excluded other study types, studies that do not meet the EPOC criteria, and studies that randomized individuals (e.g. physicians).

Types of participants

Participants were individual hospitals or groups of hospitals. We defined hospitals as healthcare institutions that have organized

medical and other professional staff and inpatient facilities, and deliver medical, nursing and related services ([WHO 2014](#)). We included all types of hospital irrespective of their specific characteristics (e.g. teaching status, ownership). We also included studies that only considered hospital units, not the whole hospital.

Types of interventions

We considered Pay-for-Performance (P4P) for inpatient health care services delivered in hospitals. We considered payments of all types of payers (public, private, insurance) at hospital level (payments on organizational level) for a defined service or a defined group of services. No other intervention features were defined as inclusion criteria (see [Description of the intervention](#)).

Types of outcome measures

Primary outcomes

P4P can be considered a complex intervention (different [interacting] components target different organizational levels and groups, flexibility of hospital response to P4P). Therefore, to allow an adequate judgement of the benefit of P4P, we considered a variety of main outcomes to weigh the advantages and disadvantages for different groups. The following outcomes were defined as main outcomes.

- Patient outcomes (quality of results): for example mortality, morbidity, patient satisfaction (e.g. Patient Satisfaction Questionnaire [PSQ] [[Grogan 2000](#)]), quality of life (e.g. EQ-5D [[Herdman 2011](#)]), adverse clinical events.
- Quality of care (quality of processes): for example adherence to recommended practice or guidelines, medication errors, readmissions.
- Utilization, coverage or access: for example waiting time, length of stay, access to service.
- Resource use, costs and cost shifting: direct medical healthcare resource use. We considered resource use irrespective of where it occurs (inpatient, outpatient, pharmacy, etc.).
- Healthcare provider outcomes: workload, work morale, stress, sick leave.
- Equity: risk selection, effects on equity for all of the other outcomes on this outcome list.
- Adverse effects or harms: adverse effects on all of the other outcomes on this outcome list.

We included only studies that reported at least one of the main outcomes.

Secondary outcomes

- Hospital volume (number of procedures per hospital): for example, admissions per year, discharges per year, procedures per year and volume shifting to other sectors (e.g. outpatient).
- Intensity (number of services per patient): for example, bed-days per patient, procedures per patient.

We planned to extract information on hospital volume and intensity, to assess if P4P affects quality directly (e.g. by quality management measures) or P4P affects quality indirectly by volume or intensity changes.

Search methods for identification of studies

Electronic searches

The EPOC Information Specialist (IS) wrote the search strategies in consultation with the authors. We searched the Cochrane Database of Systematic Reviews and the Database of Abstracts of Reviews of Effects (DARE) for related systematic reviews. We searched the following databases for primary studies on 27 June 2018.

- Cochrane Central Register of Controlled Trials (CENTRAL; 2018, Issue 5) in the Cochrane Library
- MEDLINE Ovid (including Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Versions)
- Embase Ovid
- Database of Abstracts of Reviews of Effects (DARE; 2015, Issue 2) in the Cochrane Library
- Health Technology Assessment Database (HTA; 2016, Issue 4) in the Cochrane Library

Search strategies comprise keywords and controlled vocabulary terms. We applied no language or time limits. We searched all databases from database start date to date of search. All search strategies used are provided in [Appendix 1](#).

Searching other resources

Gray Literature

We conducted a gray literature search to identify studies not indexed in the databases listed above. Sources included the sites listed below.

- Open Grey (www.opengrey.eu)
- Grey Literature Report (New York Academy of Medicine) (greyliit.org)
- Agency for Healthcare Research and Quality (AHRQ) (www.ahrq.gov)
- National Institute for Health and Care Excellence (NICE) (www.nice.org.uk)

Trial registries

We searched the following registries on 27 June 2018.

- International Clinical Trials Registry Platform (ICTRP), World Health Organization (WHO) (www.who.int/ictcp/en)
- ClinicalTrials.gov, US National Institutes of Health (NIH) (clinicaltrials.gov)

We also:

- screened individual journals and conference proceedings (e.g. via handsearching);
- reviewed the reference lists of all included studies, relevant systematic reviews/primary studies/other publications;
- contacted authors of relevant studies or reviews to clarify reported published information/seek unpublished results/data;
- contacted researchers with expertise relevant to the review topic/EPOC interventions;
- conducted cited reference searches for all included studies in citations indexes.

We also cross-checked the references of included studies and relevant systematic reviews.

Data collection and analysis

Selection of studies

Two review authors independently screened the titles and abstracts of all publications identified by the literature search and subsequently screened the full-text versions of publications for all potentially relevant titles and abstracts. We also included studies whose results were not reported in a usable manner. We resolved disagreements by discussion or by involving an independent third person. We summarized the study selection process in a PRISMA flow-chart ([Moher 2009](#)).

Data extraction and management

We extracted data into a priori piloted, standardized data extraction forms. One review author extracted descriptive data and this was verified by a second review author. Two review authors, of whom at least one was a medical statistician or epidemiologist, independently extracted data on the effects of the intervention. Disagreements were resolved by discussion. One review author entered all data into Review Manager 5 ([Review Manager 2012](#)); and, in order to avoid errors in the data entry process, all entries were checked by a second review author. We present the description and results of the included studies in 'Summary of findings' tables. We present characteristics of included studies even if they do not present usable results ([EPOC 2013c](#)). We grouped all data according to the relevant P4P program. We extracted characteristics of the analyzed hospitals (e.g. case mix), the compared payment methods and accompanying interventions (e.g. quality monitoring). We also extracted information on explanatory factors of the healthcare system (e.g. financing of the system).

Assessment of risk of bias in included studies

Two review authors independently assessed the risk of bias for each of the included studies using the 'Risk of bias' tool provided by EPOC ([EPOC 2013b](#)). We discussed discrepancies until consensus was reached.

For any included randomized, cluster randomized trials, non-randomized cluster trials, and controlled before-after studies, we assessed the following nine criteria.

- Was the allocation sequence adequately generated?
- Was the allocation adequately concealed?
- Were baseline outcome measurements similar?
- Were baseline characteristics similar?
- Were incomplete outcome data adequately addressed?
- Was knowledge of the allocated interventions adequately prevented during the study?
- Was the study adequately protected against contamination?
- Was the study free from selective outcome reporting?
- Was the study free from other risks of bias?

For any included ITS and RMS, we used the following seven criteria.

- Was the intervention independent of other changes?
- Was the shape of the intervention effect pre-specified?
- Was it unlikely that the intervention affects data collection?

- Was knowledge of the allocated interventions adequately prevented during the study?
- Were incomplete outcome data adequately addressed?
- Was the study free from selective outcome reporting?
- Was the study free from other risks of bias?

We assessed the risk of bias regarding outcome measurement under 'other source of bias' (EPOC 2013b). We judged each item to be at low, high or unclear risk of bias. For each included study, we performed the risk of bias assessment at the outcome level to assess the risk of bias for a certain outcome across studies (Guyatt 2011). If the risk of bias differed between outcomes, we performed a separate assessment for each outcome.

Measures of treatment effect

If the required data were reported in the publication, we extracted or calculated the following effect measures: For continuous variables, we used means and for dichotomous data absolute numbers and proportions. We extracted or calculated 95% confidence levels (CI) for all measures. For studies with a comparison group (randomized trials, cluster randomized trials, non-randomized cluster trials and controlled before-after studies) we extracted/calculated data for each study arm as risk ratios (RRs) for dichotomous data and as mean differences for continuous data. For CBA and ITS studies we extracted/calculated the pre- and post-intervention measures as well as the difference of the periods for specific time points. For ITS and RMS, we also extracted/calculated the pre- and post-intervention slopes of analysis, the differences of the slopes (changes in trend) and the differences of intercepts at the first intervention time point and the predicted intercept by the intervention (changes in level) (EPOC 2013d). If ITS data were inappropriately analyzed (e.g. differences of means before and after the intervention), we reanalyzed data using autoregressive integrated moving average (ARIMA) models (EPOC 2013d), if the necessary data for re-analysis could be derived from the publications. We extracted or calculated the change in difference for CBA. In case of any adjustment (e.g. for baseline) performed in the included studies, we extracted or calculated the unadjusted as well as the adjusted measures. If the data provided in the study publications could not be extracted or were not sufficiently detailed to enable recalculating as described above, we extracted data as detailed as possible.

Unit of analysis issues

The units of analysis were the hospitals, hospital units or groups of hospitals in a region, jurisdiction or in a defined healthcare system (e.g. Medicare, National Health Service (NHS), social insurance). We excluded analysis of individual physicians or other individual providers (e.g. ambulatory care facilities, rehabilitation).

Dealing with missing data

We performed all analyses without imputing missing values. If no usable numeric data were provided, we contacted the study authors to provide the necessary data.

Assessment of heterogeneity

We did not perform a meta-analysis because of the underlying clinical heterogeneity regarding hospitals, settings, design of P4P programs (see [Description of the intervention](#)), study designs, analysis methods and health system characteristics that can

influence the effect. We did not assess statistical heterogeneity because no meta-analysis was performed.

Assessment of reporting biases

We could not assess reporting bias based on asymmetry of funnel plot results, because we did not perform meta-analysis.

Data synthesis

We did not perform a meta-analysis because of the underlying heterogeneity regarding hospitals, settings, design of P4P programs (see [Description of the intervention](#)), study designs and health system characteristics that can influence the effect. We performed a structured narrative synthesis for each P4P program, considering the complexity of the intervention (Rodgers 2009; Shepperd 2009; EPOC 2017b).

Summary of findings

We used the GRADE approach to assess the certainty of effect for each outcome (Guyatt 2011). We used the worksheets for GRADE 'Summary of findings' tables (EPOC 2013c; EPOC 2017a). We summarized the findings for each intervention and graded the certainty of the evidence for each of the following most important outcomes in 'Summary of findings' tables. We judged mortality, adverse clinical events and equity as critical outcomes because these are of direct relevance to the patients and thus should also be the basis for informing policy decision making. We also judged process quality scores for measuring quality of care as an important outcome because we assumed that these are most sensitive for incentives. In addition we judged resource use as an important outcome because we presumed that this is a relevant outcome for policy decision makers due to the strong impact of the cost of hospital care on health care budgets. We prepared 'Summary of findings' tables for all outcomes judged to be critical or important. One review author graded the certainty of evidence and a second review author verified the assessment.

Subgroup analysis and investigation of heterogeneity

We aimed to analyze the influence of the following factors that can potentially moderate the relative treatment effect in our narrative synthesis. These subgroup analyses were primarily based on the within-study subgroup analyses of included studies. In addition, we considered differences of effects between studies that differed in the subgroup factors but were comparable otherwise (same P4P program, same country).

- Ownership (private hospitals/for-profit hospitals versus public/not-for-profit hospitals);
- Hospital volume/size (high volume versus low volume);
- Teaching status (teaching versus non-teaching);
- Region (urban versus rural);
- Monitoring (monitored versus not monitored);

Sensitivity analysis

We did not perform a meta-analysis and hence could not perform a quantitative sensitivity analysis. The influence on effect estimates related to differences in the risk of bias of the included studies was considered in the narrative synthesis.

RESULTS

Description of studies

Results of the search

The search of the electronic databases retrieved 10,283 studies (after electronically removing duplicates). The search of additional sources revealed one further potentially relevant study. We

screened title/abstract of these 10,284 hits. We excluded 9983 publications based on title/abstract screening. We obtained 298 full texts for detailed evaluation against the inclusion criteria. For three titles/abstracts that appeared to be potentially relevant we could not find any other data. We included 27 studies (27 publications; 20 CBA and 7 ITS) on six different P4P programs in the review. We identified one ongoing study ([Bawo 2015](#)). The process of the study selection is illustrated in the flow diagram ([Figure 1](#)).

Figure 1. Study flow diagram.

Figure 1. (Continued)

Studies
included in
quantitative
synthesis
(meta-analysis)

Included studies

Detailed descriptions of each included study can be found in the section [Characteristics of included studies](#).

Characteristics of participants, location and setting

All P4P programs targeted acute or emergency hospitals and targeted physical diseases. Four P4P programs were evaluated in the USA, one in England and one in France. The US programs encompass only the government-run Medicare or Medicaid system, or both, but not the private sector (Health Maintenance Organizations). Medicare provides health insurance for older (> 65 years) and disabled people. Medicaid provides insurance for people with limited income. The English P4P program applies to all hospitals in the National Health System (NHS), which is publicly funded and provides healthcare to all residents in England. In France the P4P program also encompassed all hospitals.

Characteristics of the interventions

[Table 1](#) provides an overview of the characteristics of the six P4P programs. The P4P programs started between 2003 and 2012. Three programs were obligatory and three programs were voluntary. All P4P programs were an add-on to capitation-based payments/diagnosis-related groups (DRGs). Two P4P program used rewards or penalties; one used first rewards and then penalties; two used penalties only and one used rewards only. Four P4P programs based their payment on quality of care and two on patient outcomes. In all P4P programs payments were made for reaching absolute quality targets. The penalty/bonus size was either fixed or depended on the annual hospital capitation budget. In three P4P programs additional payments to the payments for reaching absolute quality attainment were made for quality improvement. Payments were made either annually or ongoing (what is ongoing?). The design of the P4P programs differed widely. Therefore, we report the results for each P4P program separately.

Characteristics of comparison

In all studies the P4P program was compared to a control (e.g. region) without P4P or a time period before the implementation

of P4P. In the USA four different P4P programs were initiated. This means that in the studies on the latter P4P programs (e.g. Value-Based Purchasing Program) also hospitals in the control group may have had or have implemented other P4P programs (e.g. nonpayment for hospital-acquired conditions).

Study designs

We included 20 CBA and 7 ITS studies. All studies were funded by government agencies or received no external funding. We did not identify any randomized trials, cluster randomized trials, non-randomized cluster trials or RMS.

Outcomes

We found data on patient outcomes for four P4P programs (Premier Hospital Quality Incentive Demonstration, Value-Based Purchasing, Advancing Quality Program and Non-Payment for Hospital-Acquired Conditions) and data on quality of care for four P4P programs (Premier Hospital Quality Incentive Demonstration, Value-Based Purchasing, Financial Incentive to Quality Improvement and Hospital Readmissions Reduction). Data on equity and resource use, costs and cost shifting was only available for one P4P program (Premier Hospital Quality Incentive Demonstration).

We did not find studies reporting on utilization, coverage or access, health care provider outcomes and adverse effects or harms.

Excluded studies

Excluded studies, including exclusion reason, are listed in the section [Excluded studies](#). The list includes publications of perceived relevance (e.g. because of the title) and publications with initially discordant judgments between the review authors ([EPOC 2013c](#)).

Risk of bias in included studies

The risk of bias for each included study is presented in the 'Risk of bias' summary ([Figure 2](#)).

Figure 2. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Blinding of outcome assessment (detection bias, subjective outcomes)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias	Baseline outcomes similar	Free of contamination	Baseline characteristics similar	Intervention independent (ITS)	Shape of effect pre-specified (ITS)	Unlikely to affect data collection (ITS)	Knowledge of the allocated interventions, Blinding (ITS, objective outcomes)	Knowledge of the allocated interventions, Blinding (ITS, subjective outcomes)	Incomplete outcome data addressed (ITS)	Free of selective reporting (ITS)	Free of other bias (ITS)
Desai 2016	+	+	+	+		?	+	+	+	+	+								
Figuerola 2016	+	+	+	+		?	+	+	+	+	+								
Grossbart 2006	+	+	+	+		?	+	+	?	+	?								
Ibrahim 2017												?	+	+	+		?	+	+
Jha 2012	+	+	+	+		?	+	+	+	+	+								
Kawai 2015												?	+	+	+		?	+	+
Kristensen 2014	+	+	+	+		?	+	+	+	+	+								
Kruse 2012	+	+	+	+		?	+	+	+	+	+								
Kwong 2017	+	+	+	+		?	+	?	+	+	+								
Lalloué 2017	+	+	+	+		?	+	?	?	+	?								
Lee 2012												?	+	+	+		+	+	+
McGarry 2016	+	+	+	+		?	+	+	?	+	?								
Mellor 2017	+	+	+	+		?	+	?	?	+	?								
Morgan 2012												?	+	+	+		+	?	+
Padula 2015												?	+	+		?	?	+	+
Ryan 2009	+	+	+	+		?	+	+	+	+	+								
Ryan 2010	+	+	+	+		?	+	+	+	+	+								

Ryan 2009	+	+	+	+		?	+	+	+	+	+							
Ryan 2010	+	+	+	+		?	+	+	+	+	+							
Ryan 2011	+	+	+	+		?	+	?	+	+	?							
Ryan 2012a	+	+	+	+		?	+	+	+	+	+							
Ryan 2015	+	+	+	+	?	?	+	?	+	+	+							
Ryan 2017	+	+	+	+	?	?	+	?	+	+	+							
Schuller 2014												?	+	+	?		+	+
Shih 2014	+	+	+	+		?	+	+	?	+	+							
Sutton 2012	+	+	+	+		?	+	+	+	+	+							
Waters 2015												?	+	+	?	?	+	+
Werner 2011	+	+	+	+		?	+	+	+	+	+							
Zuckermann 2016	+	+	+	+		?	+	+	+	+	+							

compared to capitation without P4P for hospitals; [Summary of findings 6](#) Financial Incentive to Quality Improvement compared to capitation without P4P for hospitals

There were six different P4P programs.

Premier Hospital Quality Incentive Demonstration (USA)

The effect of the Premier Hospital Quality Incentive Demonstration was compared to no P4P in nine CBA ([Grossbart 2006](#); [Jha 2012](#); [Kruse 2012](#); [Ryan 2009](#); [Ryan 2010](#); [Ryan 2011](#); [Ryan 2012a](#); [Shih 2014](#); [Werner 2011](#)). The certainty of the effect estimates was downgraded because of risk of bias imprecision or both ([Summary of findings for the main comparison](#); [Table 2](#)).

The program made little or no difference on overall mortality (impact was measured in various ways). The program reduced condition-related mortality for most conditions but results were partly conflicting for different conditions and inconsistent across studies. It is uncertain whether the program impacts mortality because the certainty of this evidence is very low (5126 hospitals, 3 studies). While the program reduced adverse clinical events slightly (OR for reduction 1.11 (95% CI 0.91 to 1.36), it is uncertain whether the program impacts adverse clinical events (1139 hospitals, 1 study). The program increased process quality scores (impact was measured in various ways). It is uncertain whether the program impacts process quality scores because the certainty of this evidence is very low (4249 hospitals, 3 studies). The program demonstrated an increased inequity regarding quality of care (access to recommended care between ethnic groups), however, it is uncertain whether the program impacts inequity because the certainty of the evidence is very low (1063 hospitals, 1 study). Cost (mean hospital and costs per admission) under the program were similar to cost without P4P (impact was measured in various ways). It is uncertain whether the program has an impact on resource use because the certainty of this evidence is very low (1392 hospitals, 2 studies).

We found no study on utilization, health care provider outcomes and adverse effects.

Value-Based Purchasing (USA)

The effect of the Value-Based Purchasing Program was compared to no P4P in three CBA studies ([Figueroa 2016](#); [Ryan 2015](#); [Ryan 2017](#)). The certainty of the effect estimates was downgraded because of risk of bias and imprecision ([Summary of findings 2](#); [Table 3](#)).

Mortality was reduced and process quality scores increased in hospitals under Value-Based Purchasing (impact was measured in various ways), however, it is uncertain whether Value-Based Purchasing impacts mortality or process quality scores because the certainty of the evidence is very low (6248 hospitals, 2 studies) and (5358 hospitals, 2 studies).

We found no study on utilization, resource use, health care provider outcomes, equity and adverse effects.

One study, [Ryan 2017](#), found a small reduction for standardized patient experience measures after the introduction of P4P (difference-in-difference estimate -0.09, 95% CI -0.31 to 0.12).

Non-payment for Hospital-Acquired Conditions Program (USA)

The Non-payment for Hospital-Acquired Conditions Program was compared to no P4P in five ITS and one CBA ([Kwong 2017](#); [Lee 2012](#); [Morgan 2012](#); [Padula 2015](#); [Schuller 2014](#); [Waters 2015](#)). The certainty of the effect estimates was downgraded because of risk of bias or imprecision or both ([Summary of findings 4](#); [Table 4](#); [Table 5](#)).

Studies demonstrated fewer adverse clinical events (infections and pressure ulcers) after the introduction of the program, which used payment penalties for failure to meet quality targets (impact was measured in various ways). However, it is uncertain whether this program impacts adverse clinical events (hospital-acquired conditions) because the certainty of the evidence is very low (3568 hospitals, 6 studies).

We found no study on quality of care, resource use, health care provider outcomes, equity and adverse effects.

One study, [Morgan 2012](#), found a reduction in utilization of antimicrobials in the P4P period (change rate before 0.30; change rate after -1.24; $P < 0.001$).

Hospital Readmissions Reduction Program (USA)

The Hospital Readmissions Reduction Program was compared to no P4P in four CBA and one ITS ([Desai 2016](#); [Ibrahim 2017](#); [McGarry 2016](#); [Mellor 2017](#); [Zuckermann 2016](#)).

We found no study on patient outcomes, resource use, health care provider outcomes, equity and adverse effects.

The included studies showed no or only a very small reduction in readmissions ([Desai 2016](#); [Ibrahim 2017](#); [McGarry 2016](#); [Mellor 2017](#); [Zuckermann 2016](#); [Table 6](#); [Table 7](#)). In one CBA an increase in emergency department visits was reported (OR 1.07, 95% CI 1.04 to 1.11) and in another CBA an increase in utilization of observational services (change in regression slope 0.005, 95% CI < 0.000 to 0.009) after the implementation of P4P ([McGarry 2016](#); [Zuckermann 2016](#); [Table 6](#)). In the study of [Ibrahim 2017](#) length of stay was similar before and after the introduction of P4P (regression slope before -0.028; regression slope after -0.037).

Advancing Quality Program (UK)

The Advancing Quality Program was compared to no P4P in two CBA ([Kristensen 2014](#); [Sutton 2012](#)). The certainty of the effect estimates was downgraded because of risk of bias or imprecision, or both ([Summary of findings 5](#); [Table 8](#)).

The Advancing Quality Program reduced mortality shortly after the introduction in the reward as well as in the penalty period of the program but this reduction was not sustained (impact was measured in various ways). It is uncertain whether the program impacts mortality because the certainty of the evidence is very low (156 hospitals, 2 studies).

We found no study on quality of care, utilization, resource use, health care provider outcomes, equity and adverse effects.

Financial Incentive to Quality Improvement (France)

The effect of the Financial Incentive to Quality Improvement Program was compared to no P4P in one study ([Lalloué 2017](#); [Table 9](#)). There was a small improvement in the process quality score in

hospitals that participated in the program (difference-in-difference estimate 4.07, 95% CI -1.04 to 9.17). However, it is uncertain whether the program impacts process quality score because the certainty of this evidence is very low (377 hospitals, 1 study).

We found no study on patient outcomes, utilization, resource use, health care provider outcomes, equity and adverse effects.

Subgroup analysis (analysis of modifying design and context factors)

In three studies from the USA pre-specified subgroup analyses were performed (Lee 2012; Ryan 2012a; Ryan 2017). There is no evidence that the effect of P4P is moderated by ownership, hospital volume/size, teaching status, region or monitoring. It was not possible to analyze the effects of subgroups on the basis of differences between studies because none of these considered only a distinct subgroup (e.g. teaching hospitals).

An analysis of design factors was performed in two studies (Jha 2012; Ryan 2012a). There was no indication that the size of bonus moderated the effect. No other within-study analysis of possible effect-moderating P4P design factors was identified. Considering the impact of P4P programs across studies a larger effect size could be observed for penalties compared to bonuses (e.g. non-payment) and payments for absolute quality attainment compared to quality improvement (Table 1; Table 2 to Table 9).

DISCUSSION

Summary of main results

Effects were mostly small for all reward-based P4P programs, in particular for patient-important outcomes (mortality, adverse clinical events). These findings were broadly consistent for the different P4P programs and in different settings/contexts (across studies). Of all P4P programs, the Non-payments for Hospital-Acquired Conditions Program (a penalty-based program) showed the largest improvement of patient outcomes (clinical adverse events). However, we are uncertain whether P4P has an impact on patient outcomes because the certainty of evidence was very low.

The impact of P4P on the quality of care seems to be slightly stronger. However, it should be regarded in the interpretation of this finding that measures for quality of care (e.g. process quality scores) outcomes are at higher risk of bias because of the risk of up-coding/gaming to gain bonuses. Moreover, an improvement in quality of care (e.g. guidelines-based care) can be considered as a surrogate that might not necessarily lead to a relevant improvement of patient outcomes. It is uncertain whether P4P has an impact on quality of care because the certainty of evidence was very low.

There is only very little evidence of the impact of P4P on equity. Access to care slightly decreased under the Value-Based Purchasing Program that used rewards (or penalties) for meeting (or failure to meet) quality targets, but the impact on this equity measure is uncertain because the certainty of evidence was very low.

We could not identify any 'key' intervention components (e.g. size of incentive) that had substantial and sustained effect on the impact of P4P programs. None of our pre-specified subgroup analyses indicated that the effect is significantly modified by P4P design factors (size of incentive) or context setting (size of

hospital, region [rural vs urban], teaching status). Also no other (not prespecified by us) within-study subgroup analyses that were based on a test of interaction — including baseline hospital quality, financial situation, level of competition, participation in other quality improvement programs, public reporting, and share of private and public insured patients — had an influence. Moreover the effects did not relevantly vary between very different P4P programs. These observations suggest that the effect modification by design factors (e.g. outcome type linked to payment, frequency of payment), as well as the context/setting (e.g. baseline quality, competition) might not be strong in general. Across studies, a tendency could be observed that penalties and payments for absolute quality achievement might modify the impact of P4P. However, this should be considered as a very weak indication because it was deduced from a comparison between studies of very low certainty of evidence.

Overall completeness and applicability of evidence

All P4P programs analyzed P4P as an add-on to a capitation-based payment scheme. Consequently, the effect of P4P in addition to other basic payment schemes (e.g. global budget, fee for service) remains unclear.

We did not identify any completed studies from low- and middle-income countries and most evidence comes from P4P programs in the USA. Furthermore, all P4P programs were implemented in the public health care sector (NHS, Medicare and Medicaid, and French hospital federations under the authority of the Ministry of Health). On the one hand, the applicability of our results to private health systems and other countries might be considered limited; in particular the applicability to low- and middle-income countries should be considered with caution because of differences in health care systems and differences in coordination and organization of care. On the other hand, the results were quite similar for all countries, all P4P programs and also between different settings/contexts (e.g. medical disciplines, private versus public hospitals) suggesting that 'clinical' heterogeneity in general might not have a strong influence on the results.

Certainty of the evidence

We identified no cluster-randomized trials and all outcomes were at high risk of bias, in particular because of baseline differences between groups, contamination effects (CBA only) and possible temporal trends/temporal changes (time-varying confounding) other than the intervention. Therefore, the certainty of evidence was very low for all P4P programs and outcomes. Taking into account that the intervention effects were mostly small or very small and effect estimates were imprecise despite large sample sizes, it cannot be excluded that the observed effects are completely spurious.

Almost all studies were publicly funded, most reported wide confidence intervals and showed small effects. Therefore, we assume that there is probably no risk of reporting bias to an extent that would result in a change of our conclusion.

Potential biases in the review process

First, we might have missed some relevant studies because often P4P programs have their own labels (e.g. Value-Based Purchasing Program) or are embedded in a larger health care reform whose names do not necessarily indicate a P4P component (e.g. such as

"care act") and thus would not have been identified by the literature search unless the P4P component is indicated elsewhere in the title/abstract.

Second, there is a strong risk of overlap in our study sample (analyzed hospitals), in studies that were performed in the same country in similar time periods. The risk of a population overlap is especially high because the studies often were based on similar data sources (e.g. registries or databases).

Agreements and disagreements with other studies or reviews

As in previous systematic reviews that assessed P4P in hospitals we found no evidence for an impact of P4P on patient-important outcomes and only limited impact on process outcomes (Eijkenaar 2013; Kondo 2016; Mehrotra 2009; Mendelson 2017). In contrast to most of the other systematic reviews, we consider this finding as more uncertain because of the low quality of the underlying evidence. As in our review, the previous systematic reviews found only little information of the impact on equity aspects.

Additionally our findings on effect modifiers for the effectiveness of P4P programs in hospitals are in line with previous systematic reviews (Markovitz 2017; Van Herck 2010). None of these reviews could identify any key factors for the success of P4P programs.

AUTHORS' CONCLUSIONS

Implications for practice

We found either no difference at all or only a very slight effect for most P4P programs. For bonuses we found only short-term effects but no sustainable effects (Kristensen 2014; Sutton 2012). Penalties (e.g. non-payment for hospital-acquired conditions) seem to be slightly more effective. Considering other design factors we found that payments for quality attainment seem to be slightly more effective than payments for quality improvement. Because the certainty of evidence was very low, all these findings were uncertain. However, we think that the aspects which were the main theoretical drivers (non-randomized study designs and risk of bias) for the low certainty of evidence, would have resulted in a spurious finding in favor of P4P but not have diluted an existing effect, if they really had biased the results. Therefore, despite the low certainty of evidence we believe that these finding will probably not change largely in the future.

The findings of our review are in contrast to the logic models of the economic theory on the response of hospitals to financial incentives as well as the modifying factors (e.g. large response to larger incentives) (Barnum 1995). This shows that the results of theoretical models might not be transferable one-to-one to complex 'real world' situations. Therefore, if possible, P4P programs should be piloted under 'real life' conditions before their broad implementation. In addition, piloting has the advantage that teething problems can be identified and modified, thereby increasing the chance of success if implemented.

Our findings might indicate that hospitals have difficulties to actively modify their performance and that external factors, which are difficult to influence (e.g. availability of skilled health care professionals, hospital culture, hospital specialization), may be more important for hospital performance. Decision makers should balance the probably small long-term effects of P4P programs on patient outcomes against the costs for implementation (e.g. administrative infrastructure for measurement and reporting) and attainment (e.g. measurement of quality, allocating payments) and the fact that negative effects on equity (e.g. access to care) cannot be excluded. Moreover, the expected impact has to be judged against the impact of alternative quality measures. If decision makers nevertheless want to introduce P4P programs in hospitals, the design (e.g. size of incentive), context/setting (e.g. baseline financial situation, competition, public reporting) as well as the interaction between design factors and context/setting (e.g. necessary size of incentive if public reporting is still in effect) and context/setting factors (e.g. poor baseline performance and bad financial situation) should be carefully considered. Furthermore, possible floor and ceiling effects (e.g. very low mortality rates) should be regarded because this can influence the responsiveness of P4P. In simple terms, there must be room for improvement (e.g. low hygiene level and no hygiene standards). Moreover, the target conditions (e.g. hospital-acquired conditions) must be modifiable, i.e. there must be measures that have the potential to improve the situation (e.g. improvement of hospital hygiene). Lastly, decision makers should keep in mind that quality criteria or conditions not selected as targets for P4P might be dealt differently by hospitals than those which are relevant for P4P (Eijkenaar 2013).

Implications for research

Future studies should put a stronger focus on the P4P design factors and context/setting factors that influence the impact of P4P programs. In particular, the interaction of P4P design factors and context/setting (e.g. larger incentives in hospitals in a bad financial situation) should be evaluated. Moreover, studies in low- and middle-income countries are needed. The P4P programs and their modifying factors should be evaluated with (stepped wedge) cluster-randomized trials using established methods for the evaluation of complex interventions (Grant 2013). Future studies should use a common terminology to describe the P4P design features to facilitate their identification, analysis and synthesis.

ACKNOWLEDGEMENTS

We thank the Oxford EPOC Editorial base and the referees (Anne Lyddiat, Ndi Euphrasia Ebai-Atuh, Jonathan Fuchs, Paul Miller, Kent Ranson, Soren Kristensen, Chris Rose).

National Institute for Health Research, via Cochrane Infrastructure funding to the Effective Practice and Organisation of Care Group. The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, NHS or the Department of Health.

REFERENCES

References to studies included in this review

Desai 2016 {published data only}

Desai NR, Ross JS, Kwon JY, Herrin J, Dharmarajan K, Bernheim SM, et al. Association between hospital penalty status under the Hospital Readmission Reduction Program and readmission rates for target and nontarget conditions. *JAMA* 2016;**316**(24):2647-56.

Figueroa 2016 {published data only}

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. *BMJ* 2016;**353**:i2214.

Grossbart 2006 {published data only}

Grossbart, SR. What's the return? Assessing the effect of "pay-for-performance" initiatives on the quality of care delivery: Medical care research and review. *Medical Care Research and Review* : MCRR 2006;**63**(1_suppl):29S-48S.

Ibrahim 2017 {published data only}

Ibrahim AM, Nathan H, Thumma JR, Dimick JB. Impact of the Hospital Readmission Reduction Program on surgical readmissions among Medicare beneficiaries. *Annals of Surgery* 2017;**266**(4):617-24.

Jha 2012 {published data only}

Jha AK, Joynt KE, Orav EJ, Epstein AM. The long-term effect of premier pay for performance on patient outcomes. *New England Journal of Medicine* 2012;**366**:1606-15.

Kawai 2015 {published data only}

Kawai AT, Calderwood MS, Jin R, Soumerai SB, Vaz LE, Goldmann D, et al. Impact of the Centers for Medicare and Medicaid Services Hospital-Acquired Conditions Policy on billing rates for 2 targeted healthcare-associated infections. *Infection Control and Hospital Epidemiology* 2015;**36**(8):871-7.

Kristensen 2014 {published data only}

Kristensen SR. Long-term effect of hospital pay for performance on mortality in England. *New England Journal of Medicine* 2014;**371**:540-8.

Kruse 2012 {published data only}

Kruse GB, Polsky D, Stuart EA, Werner RM. The impact of hospital pay-for-performance on hospital and Medicare costs. *Health Services Research* 2012;**47**:2118-2136.

Kwong 2017 {published data only}

Kwong JZ, Weng Y, Finnegan M, Schaffer R, Remington A, Curtin C, et al. Effect of Medicare's nonpayment policy on surgical site infections following orthopedic procedures. *Infection Control and Hospital Epidemiology* 2017;**38**(7):817-22.

Lalloué 2017 {published data only}

Lalloué B, Jiang S, Girault A, Ferrua M, Loirat P, Minvielle E. Evaluation of the effects of the French pay-for-performance

program-IFAQ pilot study. *International Journal for Quality in Health Care* 2017;**29**(6):833-7.

Lee 2012 {published data only}

Lee GM, Kleinman K, Soumerai SB, Tse A, Cole D, Fridkin SK, et al. Effect of nonpayment for preventable infections in US hospitals. *American Journal of Infection Control* 2012;**40**(5):e190.

McGarry 2016 {published data only}

McGarry BE, Blankley AA, Li Y. The impact of the Medicare Hospital Readmission Reduction Program in New York State. *Medical Care* 2016;**54**(2):162-71.

Mellor 2017 {published data only}

Mellor J, Daly M, Smith M. Does it pay to penalize hospitals for excess readmissions? Intended and unintended consequences of Medicare's Hospital Readmissions Reductions Program. *Health Economics* 2017;**26**(8):1037-51.

Morgan 2012 {published data only}

Morgan DJ, Meddings J, Saint S, Lautenbach E, Shardell M, Anderson D, et al. Does nonpayment for hospital-acquired catheter-associated urinary tract infections lead to overtesting and increased antimicrobial prescribing?. *Clinical Infectious Diseases* 2012;**55**(7):923-9.

Padula 2015 {published data only}

Padula WV, Makic MB, Wald HL, Campbell JD, Nair KV, Mishra MK, et al. Hospital-acquired pressure ulcers at academic medical centers in the United States, 2008-2012: tracking changes since the CMS nonpayment policy. *Joint Commission Journal on Quality and Patient Safety* 2015;**41**(6):257-63.

Ryan 2009 {published data only}

Ryan AM. Effects of the Premier Hospital Quality Incentive Demonstration on Medicare patient mortality and cost. *Health Services Research* 2009;**44**(3):821-42.

Ryan 2010 {published data only}

Ryan, AM. Has pay-for-performance decreased access for minority patients?. *Health Services Research* 2010;**45**(1):6-23.

Ryan 2011 {published data only}

Ryan AM, Blustein J. The effect of the MassHealth Hospital Pay-or-Performance Program on quality. *Health Services Research* 2011;**46**(3):712-28.

Ryan 2012a {published data only}

Ryan AM, Blustein J, Doran T, Michelow MD, Casalino LP. The effect of Phase 2 of the Premier Hospital Quality Incentive Demonstration on incentive payments to hospitals caring for disadvantaged patients. *Health Services Research* 2012;**47**(4):1418-36.

Ryan 2015 {published data only}

Ryan AM, Burgess JF, Pesko MF, Borden WB, Dimick JB. The early effects of Medicare's mandatory hospital pay-for-performance program. *Health Services Research* 2015;**50**(1):81-97.

Ryan 2017 {published data only}

Ryan AM, Krinsky S, Maurer K, Dimick JB. Changes in hospital quality associated with hospital value-based purchasing. *New England Journal of Medicine* 2017;**376**(24):2358-66.

Schuller 2014 {published data only}

Schuller K, Probst J, Hardin J, Bennett K, Martin A. Initial impact of Medicare's nonpayment policy on catheter-associated urinary tract infections by hospital characteristics. *Health Policy* 2014;**115**(2-3):165-71.

Shih 2014 {published data only}

Shih T, Nicholas LH, Thumma JR, Birkmeyer JD, Dimick JB. Does pay-for-performance improve surgical outcomes? An evaluation of phase 2 of the Premier Hospital Quality Incentive Demonstration. *Annals of Surgery* 2014;**259**(4):677-81.

Sutton 2012 {published data only}

Sutton M, Nikolova S, Boaden R, Lester H, McDonald R, Roland M. Reduced mortality with hospital pay for performance in England. *New England Journal of Medicine* 2012;**367**(19):1821-8.

Waters 2015 {published data only}

Waters TM, Daniels MJ, Bazzoli GJ, Perencevich E, Dunton N, Staggs VS, et al. Effect of Medicare's nonpayment for hospital-acquired conditions: lessons for future policy. *JAMA Internal Medicine* 2015;**175**(3):347-54.

Werner 2011 {published data only}

Werner RM, Kolstad JT, Stuart EA, Polsky D. The effect of pay-for-performance in hospitals: lessons for quality improvement. *Health Affairs* 2011;**30**(4):690-8.

Zuckermann 2016 {published data only}

Zuckerman RB, Sheingold SH, Orav EJ, Ruhter J, Epstein AM. Readmissions, observation, and the Hospital Readmissions Reduction Program. *New England Journal of Medicine* 2016;**374**(16):1543-51.

References to studies excluded from this review
Atkinson 2010 {published data only}

Atkinson JG, Masiulis KE, Felgner L, Schumacher DN. Provider-initiated pay-for-performance in a clinically integrated hospital network. *Journal for Healthcare Quality* 2010;**32**(1):42-50.

Averill 2011 {published data only}

Averill RF, Hughes JS, Goldfield NI. Paying for outcomes, not performance: lessons from the Medicare Inpatient Prospective Payment System. *Joint Commission Journal on Quality and Patient Safety / Joint Commission Resources* 2011;**37**(4):184-92.

Bastian 2016 {published data only}

Bastian ND, Kang H, Nembhard HB, Bloschichak A, Griffin PM. The impact of a pay-for-performance program on central line-associated blood stream infections in Pennsylvania. *Hospital Topics* 2016;**94**(1):8-14.

Berthiaume 2006 {published data only}

Berthiaume JT, Chung RS, Ryskina KL, Walsh J, Legorreta AP. Aligning financial incentives with quality of care in the hospital setting. *Journal for Healthcare Quality* 2006;**28**(2):36-44.

Bhattacharyya 2008 {published data only}

Bhattacharyya T, Mehta P, Freiberg AA. Hospital characteristics associated with success in a pay-for-performance program in orthopaedic surgery. *Journal of Bone and Joint Surgery. American Volume* 2008;**90**(6):1240-3.

Calikoglu 2012 {published data only}

Calikoglu S, Murray R, Feeney D. Hospital pay-for-performance programs in Maryland produced strong results, including reduced hospital-acquired conditions. *Health Affairs (Project Hope)* 2012;**31**(12):2649-58.

Chen 2016 {published data only}

Chen CC, Cheng SH. Does pay-for-performance benefit patients with multiple chronic conditions? Evidence from a universal coverage health care system. *Health Policy and Planning* 2016;**31**(1):83-90.

Chen 2017 {published data only}

Chen HF, Karim S, Wan F, Nevela A, Morris ME, Bird TM, et al. Financial performance of hospitals in the Mississippi delta region under the Hospital Readmissions Reduction Program and Hospital Value-based Purchasing Program. *Medical Care* 2017;**55**(11):924-30.

Collier 2007 {published data only}

Collier, VU. Use of pay for performance in a community hospital private hospitalist group: a preliminary report. *Transactions of the American Clinical and Climatological Association* 2007;**118**:263-72.

Epstein 2014a {published data only}

Epstein AM, Jha AK, Orav EJ. The impact of pay-for-performance on quality of care for minority patients. *American Journal of Managed Care* 2014;**20**(10):e479-86.

Epstein 2014b {published data only}

Epstein AM, Joynt KE, Jha AK, Orav EJ. Access to coronary artery bypass graft surgery under pay for performance: evidence from the premier hospital quality incentive demonstration. *Circulation. Cardiovascular Quality and Outcomes* 2014;**7**(5):727-34.

Glickman 2007 {published data only}

Glickman SW, Ou FS, DeLong ER, Roe MT, Lytle BL, Mulgund J, et al. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA* 2007;**297**(21):2373-80.

Hart-Hester 2008 {published data only}

Hart-Hester S, Jones W, Watzlaf VJ, Fenton SH, Nielsen C, Madison M, et al. Impact of creating a pay for quality improvement (P4QI) incentive program on healthcare disparity: leveraging HIT in rural hospitals and small physician offices. *Perspectives in Health Information Management* 2008;**5**:1-13.

Jha 2010 {published data only}

Jha AK, Orav EJ, Epstein AM. The effect of financial incentives on hospitals that serve poor patients. *Annals of Internal Medicine* 2010;**153**(5):299-306.

Katz 2010 {published data only}

Katz, S. What is the effect of pay for performance on hospitals that serve poor patients. *Findings Brief* 2010;**13**(7):1-2.

Kim 2015 {published data only}

Kim YS, Kleerup E, Ganz PA, Ponce NA, Lorenz KA, Needleman J. Medicare payment policy creates incentives for long-term care hospitals to time discharges for maximum reimbursement. *Health Affairs (Project Hope)* 2015;**34**(6):907-15.

Kristensen 2016 {published data only}

Kristensen SR, Bech M, Lauridsen JT. Who to pay for performance? The choice of organisational level for hospital performance incentives. *European Journal of Health Economics* 2016;**17**(4):435-42.

Lindenauer 2007 {published data only}

Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, Bratzler DW. Public reporting and pay for performance in hospital quality improvement. *New England Journal of Medicine* 2007;**356**(5):486-96.

Padula 2016 {published data only}

Padula WV, Gibbons RD, Valuck RJ, Makic MB, Mishra MK, Pronovost PJ, Meltzer DO. Are evidence-based practices associated with effective prevention of hospital-acquired pressure ulcers in US academic medical centers?. *Medical Care* 2016;**54**(5):512-8.

Park 2011 {published data only}

Park MH, Hiller EA. Medicare Hospital Value-Based Purchasing: the evolution toward linking Medicare reimbursement to health care quality continues. *Health Care Law Monthly* 2011;**2**:2-9.

Rieger 2009 {published data only}

Rieger UM, Prengel A, Burla S, Rüdiger M, Pierer G, Heberer M. From pay-for-effort to pay-for-performance – an analysis of the Swiss health care system with focus on the inpatient sector. *Praxis* 2009;**98**(25):1499-1509.

Ryan 2012b {published data only}

Ryan AM, Blustein J, Doran T, Michelow MD, Casalino LP. The effect of Phase 2 of the Premier Hospital Quality Incentive Demonstration on incentive payments to hospitals caring for disadvantaged patients. *Health Services Research* 2012;**47**(4):1418-36.

Ryan 2014 {published data only}

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 Services Research* 2014;**49**(2):568-87.

Sautter 2007 {published data only}

Sautter KM, Bokhour BG, White B, Young GJ, Burgess JF Jr, Berlowitz D, Wheeler JR. The early experience of a hospital-

based pay-for-performance program. *Journal of Healthcare Management / American College of Healthcare Executives* 2007;**52**(2):95-107.

Thirukumaran 2017 {published data only}

Thirukumaran CP, Glance LG, Temkin-Greener H, Rosenthal MB, Li Y. Impact of Medicare's Nonpayment Program on hospital-acquired conditions. *Medical Care* 2017;**55**(5):447-55.

Vaz 2015 {published data only}

Vaz LE, Kleinman KP, Kawai AT, Jin R, Kassler WJ, Grant PS, et al. Impact of Medicare's hospital-acquired condition policy on infections in safety net and non-safety net hospitals. *Infection Control and Hospital Epidemiology* 2015;**36**(6):649-55.

References to ongoing studies
Bawo 2015 {published data only}

Quality-based pay for performance scheme in Liberia. Ongoing study Unclear.

Additional references
Abel-Smith 1994

Abel-Smith B, Mossialos E. Cost containment and health care reform: a study of the European Union. *Health Policy* 1994;**28**(2):89-132.

Barnum 1995

Barnum H, Kutzin J, Saxenian H. Incentives and provider payment methods. *International Journal of Health Planning and Management* 1995;**10**(1):23-45.

Blomqvist 1991

Blomqvist Å. The doctor as double agent: information asymmetry, health insurance, and medical care. *Journal of Health Economics* 1991;**10**(4):411-32.

Böcking 2005

Böcking W, Ahrens U, Kirch W, Milakovic M. First results of the introduction of DRGs in Germany and overview of experience from other DRG countries. *Journal of Public Health* 2005;**13**(3):128-37.

Campbell 2007

Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. *BMJ* 2007;**334**(7591):455-9.

Chaix-Couturier 2000

Chaix-Couturier C, Durand-Zaleski I, Jolly D, Durieux P. Effects of financial incentives on medical practice: results from a systematic review of the literature and methodological issues. *International Journal for Quality in Health Care* 2000;**12**(2):133-42. [PUBMED: 10830670]

Craig 2008

Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;**337**:a1655.

Dixon 2004

Dixon J. Payment by results - new financial flows in the NHS. *BMJ* 2004;**328**(7446):969-70.

Draper 2006

Draper D, Kahn KL, Reinisch EJ, Sherwood MJ, Carney MF, Kosecoff J, et al. Effects of Medicare's prospective payment system on the quality of hospital care. www.rand.org/pubs/research_briefs/RB4519-1/index1.html (accessed prior to 29 May 2019).

Eijkenaar 2013

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**(2-3):115-30.

Ellis 1986

Ellis RP, McGuire TG. Provider behavior under prospective reimbursement. Cost sharing and supply. *Journal of Health Economics* 1986;**5**(2):129-51.

Ellis 1996

Ellis RP, McGuire TG. Hospital response to prospective payment: moral hazard, selection, and practice-style effects. *Journal of Health Economics* 1996;**15**(3):257-77.

EPOC 2013a

Cochrane Effective Practice, Organisation of Care (EPOC). What study designs should be included in an EPOC review and what should they be called?. epoc.cochrane.org/epoc-author-resources (accessed 26 May 2014).

EPOC 2013b

Cochrane Effective Practice, Organisation of Care (EPOC). Suggested risk of bias criteria for EPOC reviews. epoc.cochrane.org/epoc-author-resources (accessed 26 May 2014).

EPOC 2013c

Cochrane Effective Practice, Organisation of Care (EPOC). Analysis in EPOC reviews. epoc.cochrane.org/epoc-author-resources (accessed 26 May 2014).

EPOC 2013d

Cochrane Effective Practice, Organisation of Care (EPOC). Interrupted time series (ITS) analyses. epoc.cochrane.org/epoc-author-resources (accessed 26 May 2014).

EPOC 2017a

Cochrane Effective Practice, Organisation of Care (EPOC). EPOC Worksheets for preparing a 'Summary of findings' table using GRADE.. epoc.cochrane.org/resources/epoc-resources-review-authors (accessed 29 September 2017).

EPOC 2017b

Cochrane Effective Practice, Organisation of Care (EPOC). Synthesising results when it does not make sense to do a meta-analysis. epoc.cochrane.org/resources/epoc-resources-review-authors (accessed 29 September 2017).

Epstein 2012

Epstein AM. Will pay for performance improve quality of care? The answer is in the details. *New England Journal of Medicine* 2012;**367**(19):1852-3. [DOI: [10.1056/NEJMe1212133](https://doi.org/10.1056/NEJMe1212133)]

Gosden 2000

Gosden T, Forland F, Kristiansen I, Sutton M, Leese B, Giuffrida A, et al. Capitation, salary, fee-for-service and mixed systems of payment: effects on the behaviour of primary care physicians. *Cochrane Database of Systematic Reviews* 2000, Issue 3. [DOI: [10.1002/14651858.CD002215](https://doi.org/10.1002/14651858.CD002215)]

Grant 2013

Grant A, Treweek S, Dreischulte T, Foy R, Guthrie B. Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting. *Trials* 2013;**14**(1):15.

Grogan 2000

Grogan S, Conner M, Norman P, Willits D, Porter I. Validation of a questionnaire measuring patient satisfaction with general practitioner services. *Quality in Health Care* 2000;**9**(4):210-5.

Guyatt 2011

Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology* 2011;**64**(4):383-94. [DOI: [10.1016/j.jclinepi.2010.04.026](https://doi.org/10.1016/j.jclinepi.2010.04.026)]

Herdman 2011

Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research* 2011;**20**(10):1727-36.

Jon 2012

Jon B, Douglas C. Provider payment and incentives. In: Glied S, Smith PC editor(s). *The Oxford Handbook of Health Economics*. Oxford: Oxford University Press, 2012:624-48.

Kondo 2016

Kondo KK, Damberg CL, Mendelson A, Motu'apuaka M, Freeman M, O'Neil M, et al. Implementation processes and pay for performance in healthcare: a systematic review. *Journal of General Internal Medicine* 2016;**31**(Suppl 1):61-9.

Ma 1994

Ma C-TA. Health care payment systems: cost and quality incentives. *Journal of Economics & Management Strategy* 1994;**3**(1):93-112.

Markovitz 2017

Markovitz AA, Ryan AM. Pay-for-performance: disappointing results or masked heterogeneity?. *Medical Care Research and Review* : *MCRR* 2017;**74**(1):3-78.

Mehrotra 2009

Mehrotra A, Damberg CL, Sorbero ME, Teleki SS. Pay for performance in the hospital setting: what is the state of the evidence?. *American Journal of Medical Quality* 2009;**24**(1):19-28.

Mendelson 2017

Mendelson A, Kondo K, Damberg C, Low A, Motúapuaka M, Freeman M, O'Neil M, Relevo R, Kansagara D. The effects of pay-for-performance programs on health, health care use, and processes of care: a systematic review. *Annals of Internal Medicine* 2017;**166**(5):341-53.

Moher 2009

Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine* 2009;**151**(4):264-9.

Petersen 2006

Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care?. *Annals of Internal Medicine* 2006;**145**(4):265-72.

Petticrew 2013

Petticrew M, Anderson L, Elder R, Grimshaw J, Hopkins D, Hahn R, Krause L, Kristjansson E, Mercer S, Sipe T, Tugwell P, Ueffing E, Waters E, Welch V. Complex interventions and their implications for systematic reviews: a pragmatic approach. *Journal of Clinical Epidemiology* 2013;**66**(11):1209-1214.

Porter 2010

Porter ME. What is value in health care?. *New England Journal of Medicine* 2010;**363**:2477-81.

Review Manager 2012 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2012.

Rodgers 2009

Rodgers M, Sowden A, Petticrew M, Arai L, Roberts H, Britten N, et al. Testing methodological guidance on the conduct of narrative synthesis in systematic reviews: effectiveness of interventions to promote smoke alarm ownership and function. *Evaluation* 2009;**15**(1):49-73.

Shepperd 2009

Shepperd S, Lewin S, Straus S, Clarke M, Eccles MP, Fitzpatrick R, et al. Can we systematically review studies that evaluate complex interventions?. *PLoS Medicine* 2009;**6**(8):1-8.

Trochim 2006

Trochim WM, Cabrera DA, Milstein B, Gallagher RS, Leischow SJ. Practical challenges of systems thinking and modelling in public health. *American Journal of Public Health* 2006;**96**(3):538-46.

Van Herck 2010

Van Herck P, De Smedt D, Annemans L, Remmen R, Rosenthal MB, Sermeus W. Systematic review: effects, design choices, and context of pay-for-performance in health care. *BMC Health Services Research* 2010;**10**(1):247.

WHO 2013

World Health Organization (WHO). How are hospitals funded and which payment method is best?. <http://www.euro.who.int/en/data-and-evidence/evidence-informed-policy-making/publications/hen-summaries-of-network-members-reports/how-are-hospitals-funded-and-which-payment-method-is-best> (accessed 26 May 2014).

WHO 2014

World Health Organization (WHO). Hospitals. www.who.int/topics/hospitals/en/ (accessed 26 May 2014).

Witter 2012

Witter S, Fretheim A, Kessy FL, Lindahl AK. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database of Systematic Reviews* 2012, issue 2. [PUBMED: 10.1002/14651858.CD007899.pub2]

Zweifel 2009

Zweifel P, Breyer F, Kifmann M. Paying providers. In: Zweifel P, Breyer F, Kifmann M editor(s). *Health Economics*. Berlin: Springer, 2009:331-77.

References to other published versions of this review

Mathes 2014

Mathes T, Pieper D, Mosch CG, Jaschinski T, Eikermann. Payment methods for hospitals. *Cochrane Database of Systematic Reviews* 2014, Issue 6. [DOI: [10.1002/14651858.CD011156](https://doi.org/10.1002/14651858.CD011156)]

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Desai 2016

Methods

Study design (CBA, ITS): CBA

Allocation linked to study or natural experiment: natural experiment

Selection of hospitals (intervention/control):

- Penalized hospitals (target vs. non-target conditions)

Desai 2016 (Continued)

- Non-penalized hospitals (target vs. non-target conditions)

Unit of allocation (region, hospital, country): hospitals

Nature of desired change: implementation

Data source: Medicare fee-for-service claims data for 1 January 2008, through 30 June 2015, to identify hospital admissions. Data on which hospitals were subject to penalties at the time the HRRP was implemented in October 2012 from the Centers for Medicare & Medicaid Services website. For condition-specific measures, we used *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)* codes to identify discharges of Medicare beneficiaries

Unit of analyses: hospitals

Number of measurements (before, transition, after, unit [e.g. years]): 27/30/33 (months)

Statistical analyses:

- Method: difference-interrupted-time-series models with ordinary linear regression.
- Adjustment factors: not specified ("all risk factors from the corresponding publicly reported measure")

Participants

Country/region/setting: USA/whole country/hospitals

Health system characteristics: Medicare & Medicaid

Number of hospitals included in the analysis (intervention/control): 2214/1283

Characteristics of hospitals (intervention/control):

- Number of beds:
 - * 6 to 99: 571 (25.8%)/546 (42.6%)
 - * 100 to 199: 633 (28.6%)/278 (21.7%)
 - * 200 to 299: 380 (17.2%)/161 (12.6%)
 - * 300 to 399: 222 (10.0%)/97 (7.6%)
 - * 400 to 499: 126 (5.7%)/52 (4.1%)
 - * ≥ 500: 206 (9.3%)/68 (5.3%)
- Missing data: 76 (3.4)/81 (6.3)
- Number of patients: NR
- Number of wards: NR
- Departments: NR
- Owner: NR
 - * public: 330 (14.9%)/186 (14.5%)
 - * not for profit: 1321 (59.7%)/708 (55.2%)
 - * for profit: 487 (22.0%)/308 (24.0%)
 - * missing data: 76 (3.4)/81 (6.3)
- Teaching status:
 - * teaching: 732 (33.1%)/376 (29.3%)
 - * nonteaching: 1406 (63.5%)/826 (64.4%)
- Location:
 - * West: 332 (15.0)/297 (23.2)
 - * Midwest: 464 (21.0)/294 (22.9)
 - * Northeast: 405 (18.3)/97 (7.6)
 - * South: 937 (42.3)/467 (36.4)
 - * associated areas: 0 (0)/47 (3.7)
 - * missing data: 76 (3.4)/81 (6.3)

Desai 2016 (Continued)

- Setting:
 - * urban: 1927 (87.0%)/1103 (86.0%)
 - * rural: 211 (9.5%)/99 (7.7%)

Number of patients included in the analysis (total): 20,351,161

Characteristics of patients (before [whole population], intervention/control):

- Age: ≥ 65
- Gender: NR
- Casemix: NR
- Indications: acute myocardial infarction, congestive heart failure, and pneumonia

Existing/other quality programs: NR

Other relevant context information: none

Interventions	Penalties beginning in October 2012 for hospitals with higher than expected readmissions for acute myocardial infarction, congestive heart failure, and pneumonia among their fee-for-service Medicare beneficiaries. Since the program's inception, thousands of hospitals have been subjected to penalties now totaling nearly USD 1 billion.
	<p>Control:</p> <p>No penalties</p>
Outcomes	30-day, risk-adjusted, all-cause unplanned readmission
Notes	<p>Funding/conflict of interest: study was funded by the Agency for Healthcare Research and Quality. Dr Desai is supported by grant from the Agency for Healthcare Research and Quality. Dr Dharmarajan is supported by grant from the National Institute on Aging and the American Federation for Aging Research through the Paul B. Beeson Career Development Award Program. He is also supported by grant P30AG021342 via the Yale Claude D. Pepper Older Americans Independence Center.</p> <p>Information for subgroup analysis: -</p> <p>Other comments: -</p>

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention cannot be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Readmission
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not reported

Pay for performance for hospitals (Review)

Desai 2016 (Continued)

Selective reporting (re-reporting bias)	Low risk	No indication for selective reporting
Other bias	High risk	Analysed outcome same as outcome linked to payment
Baseline outcomes similar All outcomes	High risk	Baseline outcome measures not similar
Free of contamination	High risk	Allocation on individual level
Baseline characteristics similar	High risk	Baseline characteristics not similar

Figueroa 2016

Methods	<p>Study design (CBA, ITS): CBA</p> <p>Allocation linked to study or natural experiment: natural experiment</p> <p>Selection of hospitals (intervention/control): comparison 1: all conditions targeted by P4P/selected conditions not targeted by P4P; comparison 2: acute care hospitals/hospitals in another region and critical access hospitals</p> <p>Unit of allocation (region, hospital, country): comparison 1: individual level; comparison 2: region and hospital type</p> <p>Nature of desired change: introduction</p> <p>Data source: 100% Medicare inpatient claims data from 2008 through 2013</p> <p>Unit of analyses: comparison 1: individual; comparison 2: hospital</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): 14/NA/10 (quarter)</p> <p>Statistical analyses:</p> <ul style="list-style-type: none"> - Method: difference-in-difference; random effects linear spline regression - Adjustment factors: comorbidities, seasonal variation
Participants	<p>Country/region/setting: USA/whole country/hospitals</p> <p>Health system characteristics: Medicare</p> <p>Number of hospitals included in the analysis (intervention/control): 2919/1348</p> <p>Characteristics of hospitals (before [whole population], intervention/control): acute care hospitals/hospitals in another region and critical access hospitals</p> <ul style="list-style-type: none"> • Number of beds: <ul style="list-style-type: none"> * small: 27.7%/92.7% * medium: 57.4%/6.8% * large: 13.9%/0.6% • Number of patients (mean annual Medicare volume): 2671/385 • Number of wards: NR • Departments: NR

Figueroa 2016 (Continued)

- Owner:
 - * For profit: 20.4%/5.3%
 - * Private not for profit: 64.8%/54.5%
 - * Public: 14.8%/40.3%
- Teaching status:
 - * major: 9.0%/0.5%
 - * minor: 24.0%/6.4%
 - * none: 67.1%/93.2%
- Location: NR

Number of patients included in the analysis (intervention/control): 2,252,818/177,800

Characteristics of patients (intervention/control): acute

myocardial infarction, congestive heart failure, and pneumonia/stroke, sepsis, gastroenteritis and esophagitis, gastrointestinal bleed, urinary tract infection, metabolic disorder, arrhythmia, renal failure

- Age (mean): 79.8/81.0
- Gender (male): 42.1%/39.5%
- Casemix: NR
- Indications: NR

Existing/other quality programs: not reported

Other relevant context information: none

Interventions	<p>Rewards or penalizes hospitals based on their performance on multiple domains of care, including clinical processes, clinical outcomes (e.g. 30-day mortality for acute myocardial infarction, pneumonia, and heart failure), patient experience, and, latter, cost efficiency.</p> <p>Performance is determined based on hospitals' absolute achievement compared with the national average, or improvement compared with their own performance in the baseline period, depending on which is greater.</p> <p>Funding is designed to be budget neutral; Medicare withholds a percentage of inpatient payments to prospectively paid hospitals and then redistributes this money back to hospitals based on their performance.</p> <p>National in scope and obligatory.</p> <p>Control:</p> <p>Comparison 1: No P4P</p> <p>Comparison 2: Probably mixture of basic P4P (incentives of hospitals without penalties) and no P4P</p>
Outcomes	Mortality (30 days)
Notes	<p>Funding/conflict of interest: work received no support from any organization; authors no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years; no other relationships or activities that could appear to have influenced the submitted work</p> <p>Information for subgroup analysis: low baseline performance</p> <p>Other comments: -</p>
Risk of bias	
Bias	Authors' judgement Support for judgement

Figueroa 2016 (Continued)

Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention cannot be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Mortality
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not sufficiently reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Low risk	No evidence of other risk of bias
Baseline outcomes similar All outcomes	High risk	Difference in baseline outcomes
Free of contamination	High risk	Comparison 1: high risk, allocation on individual level and intervention cannot be blinded Comparison 2: low risk, allocation on hospital level
Baseline characteristics similar	High risk	Difference in baseline characteristics

Grossbart 2006

Methods	<p>Study design (CBA, ITS): CBA</p> <p>Allocation linked to study or natural experiment: linked to study</p> <p>Selection of hospitals (intervention/control): Nonrandom sample of hospitals that were eligible to participate in the Centers for Medicare & Medicaid Services/Premier HQID Project. A test group of 4 acute care hospitals within Catholic Healthcare Partners that are participating in this demonstration project was compared with a control group of 6 hospitals in the same health care system that chose not to participate in the project. To ensure a level of homogeneity among the hospitals in this study, analysis was limited to hospitals with similar levels of service.</p> <p>Unit of allocation (region, hospital, country): hospitals</p> <p>Nature of desired change: initiation of P4P</p> <p>Data source: Data for this analysis were obtained from Catholic Healthcare Partners' quality measures database that is provided by the system's core measures vendor</p> <p>Unit of analyses: hospitals</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): 1/0/1 (years)</p>
---------	---

Pay for performance for hospitals (Review)

Grossbart 2006 (Continued)

Statistical analyses:

- Method: z-Test
- Adjustment factors: none

Participants

Country/region/setting: USA/5 states/acute care hospitals

Health system characteristics: 9 regional service areas in 5 states and includes 29 hospitals, several long-term care facilities, housing sites for the elderly, home health agencies, hospice programs, out-reach services, medical groups, wellness centers, and other organizations that operate diversified health care activities.

Medicare and Medicaid patients

Number of hospitals included in the analysis (intervention/control): 4/6

Characteristics of hospitals (intervention/control):

- Number of beds (range): 158 to 422/141 to 463
- Number of patients (admissions per year, range): 8541 to 19,225/8936 to 21,634
- Number of wards: NR
- Departments: NR
- Owner: non-profit (catholic)
- Teaching status: NR
- Location: NR

Number of patients included in the analysis (before/after or intervention/control): NR

Characteristics of patients (intervention/control):

- Age: NR
- Gender: NR
- Casemix (range): 1.15 to 1.71/1.26 to 1.46
- Indications: acute myocardial infarction, coronary artery bypass graft, pneumonia

Existing/other quality programs: public quality reporting

Other relevant context information: none

Interventions

Financial and other incentives based on 35 quality measures in 5 clinical areas: acute myocardial infarction, heart failure, pneumonia, coronary artery bypass graft, and joint replacement of the hip or knee. For each clinical area in the project, hospitals with a composite quality score in the top 10% of participants received a 2% incentive bonus on top of Medicare payment for traditional fee-for-service patients within that specific clinical condition. Hospitals in the second decile received a 1% incentive bonus, while those performing above the median composite quality score was publicized as top performers by Centers for Medicare & Medicaid Services. The project also includes a slight downside risk in its 3rd year for low performers that fail to rise above threshold quality scores set in the 1st year at the lowest 2 deciles.

Outcomes
Quality composite score (percentage of complied criteria)

- Acute myocardial infarction: (1) aspirin at arrival, (2) aspirin at discharge, (3) angiotensin converting enzyme inhibitor (ACEI) for left ventricular systolic dysfunction, (4) Smoking cessation device/counseling, (5) beta blocker at arrival, (6) beta blocker at discharge, (7) thrombolytic agent within 30 minutes of arrival, (8) percutaneous coronary intervention within 120 minutes of arrival.
- Heart failure: (1) left ventricular function assessment, (2) detailed discharge instructions, (3) angiotensin converting enzyme inhibitor for left ventricular systolic dysfunction, (4) smoking cessation advice/counseling.
- Pneumonia: (1) percentage of patients who received an oxygenation assessment within 24 hours prior to, or after, hospital arrival; (2) blood culture collected prior to first antibiotic administration; (3) pneumococcal screening/vaccination; (4) antibiotic timing, percentage of pneumonia patients who re-

Grossbart 2006 (Continued)

ceived first dose of antibiotics within 4 hours after hospital arrival; (5) smoking cessation advice/counseling.

Notes

Funding/conflict of interest: NR

Information for subgroup analysis: NR

Other comments: NR

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention cannot be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Quality composite score
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Only complete cases were included. Proportion missing for each group not reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Low risk	No evidence for other risk of bias
Baseline outcomes similar All outcomes	Unclear risk	Not reported
Free of contamination	Low risk	It is unlikely that the control received the intervention
Baseline characteristics similar	Unclear risk	Not reported

Ibrahim 2017

Methods

Study design (CBA, ITS): ITS

Allocation linked to study or natural experiment: natural experiment

Selection of hospitals (intervention/control): penalized hospitals. Hospitals included in this study were identified by their provider number in the Medicare Provider Analysis and Review file

Unit of allocation (region, hospital, country): NA

Nature of desired change: implementation

Ibrahim 2017 (Continued)

Data source: We used data from the Medicare Provider Analysis and Review file including procedures from 2008 to 2014. *International Classification of Disease—Clinical Modification, 9th Edition* codes were used to identify a total of 8 different surgical procedures

Unit of analyses: hospitals

Number of measurements (before, transition, after, unit [e.g. years]): 9/10/9 (quarters)

Statistical analyses:

- Method: interrupted time series
- Adjustment factors: NR

Participants	<p>Country/region/setting: USA/whole country/hospitals</p> <p>Health system characteristics: Medicare & Medicaid</p> <p>Number of hospitals included in the analysis (before): 3497</p> <p>Characteristics of hospitals (before):</p> <ul style="list-style-type: none"> • Number of beds: <ul style="list-style-type: none"> * < 250 beds: 37.2% * 250 to < 500 beds: 37.4% * ≥ 500 beds: 25.5% • Number of patients: NR • Number of wards: NR • Departments: NR • Owner: <ul style="list-style-type: none"> * for-profit: 14.4% * nonprofit: 77.1% * other: 8.5% • Teaching status: <ul style="list-style-type: none"> * yes: 60.0 <p>Number of patients included in the analysis (before/after or intervention/control): NR</p> <p>Characteristics of patients (before): 5,122,240</p> <ul style="list-style-type: none"> • Age (mean): 74.5 • Gender (men): 43.3% • Casemix: NR • Indications: total hip replacement, total knee replacements <p>Existing/other quality programs: NR</p> <p>Other relevant context information: none</p>
Interventions	<p>Under this policy, hospitals with higher than expected readmissions rates would be subject to payment penalties</p> <p>Control: no penalties</p>
Outcomes	Readmission (30 days), length of stay
Notes	<p>Funding/conflict of interest: AMI acknowledges funding from the Robert Wood Johnson Foundation and the United States Department of Veterans Affairs supporting his role as a Robert Wood Johnson Clinical Scholar. HN acknowledges funding from the Agency for Healthcare Research and Quality under award number K08HS024763-01. JBD acknowledges funding from the National Institute of Aging of the National Institute of Health under award number R01AG039434-04. JBD has a financial interest in Ar-</p>

Ibrahim 2017 (Continued)

borMetrix, Inc., which had no role in the analysis herein. The remaining authors have no conflicts of interest to disclose.

Information for subgroup analysis: NR

Other comments: none

Risk of bias

Bias	Authors' judgement	Support for judgement
Intervention independent (ITS)	Unclear risk	No information
Shape of effect pre-specified (ITS)	Low risk	Shape pre-specified
Unlikely to affect data collection (ITS) All outcomes	High risk	Analysed outcome same as outcome linked to payment
Knowledge of the allocated interventions, Blinding (ITS, objective outcomes)	Low risk	Readmission objective outcome
Incomplete outcome data addressed (ITS) All outcomes	Unclear risk	No information on missing outcome data
Free of selective reporting (ITS)	High risk	Focus on statistical significance
Free of other bias (ITS)	Low risk	No evidence for other bias

Jha 2012
Methods

Study design (CBA, ITS): CBA

Allocation linked to study or natural experiment: linked to study

Selection of hospitals (intervention/control):

- Intervention hospitals in the Premier Healthcare Informatics program were invited by CMS to participate in the HQID, of which 60% joined and were available for analysis.
- Control group: all non-Premier hospitals that reported to Hospital Compare

Unit of allocation (region, hospital, country): hospital

Nature of desired change: initiation of P4P

Data source: national Medicare Part A data

Unit of analyses: hospitals

Number of measurements (before, transition, after, unit [e.g. years]): 6/0/26 (quarters)

Statistical analyses:

- Method: linear regression

Jha 2012 (Continued)

- Adjustment factors: size, teaching status, location, ownership, region, financial margin, the Herfindahl–Hirschman index, proportion of patients receiving Medicare

Participants	<p>Country/region/setting: USA/all/hospitals</p> <p>Health system characteristics: Medicare and Medicaid patients</p> <p>Number of hospitals included in the analysis (intervention/control): 252/3363</p> <p>Characteristics of hospitals (intervention/control):</p> <p>Number of beds:</p> <ul style="list-style-type: none"> * < 100: 13.49%/39.77% * 100 to < 400: 61.11%/50.13% * ≥ 400: 25.40%/11.09% <ul style="list-style-type: none"> • Number of patients: 137,287/1,069,034 • Number of wards: NR • Departments: NR • Owner: <ul style="list-style-type: none"> * private for-profit: 1.19%/17.78% * private nonprofit: 90.08%/61.64% * public: 8.73%/20.58% • Teaching status: <ul style="list-style-type: none"> * teaching: 13.49%/7.11% * nonteaching: 86.51%/92.89% • Location: <ul style="list-style-type: none"> * urban: 94.84%/81.56% * rural: 5.16%/18.44% <p>Number of patients included in the analysis (before/after or intervention/control): NR</p> <p>Characteristics of patients (intervention/control):</p> <ul style="list-style-type: none"> • Age (mean): 79.92/79.63 • Gender (female): 51.49%/52.52% • Casemix: NR • Indications: acute myocardial infarction, congestive heart failure, pneumonia, and coronary artery bypass <p>Existing/other quality programs: public reporting</p> <p>Other relevant context information: none</p>
Interventions	Hospitals that performed in the top 2 deciles for selected conditions were eligible for 1% to 2% bonuses in Medicare payments for that condition, whereas under performing hospitals were liable for a 1% to 2% financial penalty starting in the fourth year of the program. The Premier HQID made modest changes later in the program to offer additional incentives for hospitals that made substantial improvements in care.
Outcomes	Mortality (30 days)
Notes	<p>Funding/conflict of interest: Supported by a grant from the Robert Wood Johnson Foundation, all authors disclosed grant to their institution.</p> <p>Information for subgroup analysis: -</p> <p>Other comments: -</p>

Jha 2012 (Continued)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention cannot be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Mortality, objective outcome
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Low risk	No evidence for other risk of bias
Baseline outcomes similar All outcomes	High risk	Outcome measurements differ
Free of contamination	Low risk	It is unlikely that the control received the intervention
Baseline characteristics similar	Low risk	Adjusted for in the analysis

Kawai 2015

Methods

Study design (CBA, ITS): ITS

Allocation linked to study or natural experiment: natural experiment

Selection of hospitals (intervention/control): Eligible hospitals were those subject to the Inpatient Prospective Payment System. Federal, critical access, long-term care, cancer, psychiatric, children's, and rehabilitation hospitals were excluded, since they were not subject to the Inpatient Prospective Payment System or hospital-acquired conditions payment policies.

Unit of allocation (region, hospital, country): NA

Nature of desired change: implementation

Data source: Billing rates for healthcare-associated infections data were obtained from the State Inpatient Databases, Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. Billing for vascular catheter-associated infections and catheter-associated urinary tract infections was ascertained in claims data using CMS HAC definitions, which incorporate ICD-9 codes and present-on-admission indicators. Hospital characteristics were obtained by linking the State Inpatient Databases to the 2009 American Hospital Association Annual Survey Database using hospital identifiers

Kawai 2015 (Continued)

Unit of analyses: hospitals

Number of measurements (before, transition, after, unit [e.g. years]): 3/12 for VCAI; 7/12 for CAUTI (quarter)

Statistical analyses:

- Method: logistic regression mixed-effects models clustered by hospital

- Adjustment factors: state, hospital size, ownership type, teaching status and percent Medicare admissions

Participants

Country/region/setting: USA/California, Massachusetts, New York/acute care hospitals

Health system characteristics: Medicare

Number of hospitals included in the analysis (before): 569

Characteristics of hospitals (before):

- Number of beds:
 - * ≤ 99: 20%
 - * 100 to 399: 65%
 - * ≥400: 15%
- Number of patients: NR
- Number of wards: NR
- Departments: NR
- Owner:
 - * public: 14%
 - * for-profit: 15%
 - * not-for-profit: 71%
- Teaching status:
 - * major: 12%
 - * graduate: 17%
 - * limited: 8%
 - * nonteaching: 63%
- Location:
 - * metropolitan: 92%
 - * micropolitan: 7%
 - * rural: 2%

Number of patients included in the analysis (before): 24,298 discharges (vascular catheter-associated infections), 38,326 discharges (catheter-associated urinary tract infections)

Characteristics of patients (before [whole population], intervention/control):

- Age: NR
- Gender: NR
- Casemix: NR
- Indications: NR

Existing/other quality programs: NR

Other relevant context information: none

Interventions

Limited additional payment for selected hospital-acquired conditions considered “reasonably preventable”. Hospital-acquired conditions include 2 healthcare-associated infections: vascular catheter-associated infections and catheter-associated urinary tract infections.

Kawai 2015 (Continued)

Control: no penalties

Outcomes	Hospital acquired conditions
Notes	<p>Funding/conflict of interest: Financial support by the Agency for Healthcare Research and Quality (grant R01HS018414). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality. All authors report no conflicts of interest relevant to this article.</p> <p>Information for subgroup analysis: owner, size, proportion of Medicare admissions</p> <p>Other comments: -</p>

Risk of bias

Bias	Authors' judgement	Support for judgement
Intervention independent (ITS)	Unclear risk	No information
Shape of effect pre-specified (ITS)	Low risk	Shape pre-specified
Unlikely to affect data collection (ITS) All outcomes	High risk	Analysed outcome same as outcome linked to payment
Knowledge of the allocated interventions, Blinding (ITS, objective outcomes)	Low risk	Infections
Incomplete outcome data addressed (ITS) All outcomes	Unclear risk	No information
Free of selective reporting (ITS)	Low risk	No evidence for selective reporting
Free of other bias (ITS)	Low risk	No evidence for other bias

Kristensen 2014

Methods	<p>Study design (CBA, ITS): CBA</p> <p>Allocation linked to study or natural experiment: natural experiment</p> <p>Selection of hospitals (intervention/control): all hospitals providing emergency care in the north-west region of England participated in the Advancing Quality program/rest of England</p> <p>Unit of allocation (region, hospital, country): region</p> <p>Nature of desired change: initiation of P4P</p> <p>Data source: quality measures related to the incentive program were obtained from Advancing Quality administrators. Data on patient characteristics, coexisting conditions, and mortality were obtained from national Hospital Episode Statistics</p> <p>Unit of analyses: hospitals</p>
---------	--

Kristensen 2014 (Continued)

Number of measurements (before , transition, after, unit [e.g. years]): 6/0/6 (period 1), 8 (period 2) (quarter)

Statistical analyses:

- Method: difference-in-difference (logistic regression)
- Adjustment factors: sex and age, 31 coexisting conditions, type of admission (emergency or transfer from another hospital), and the location from which the patient was admitted

Participants

Country/region/setting: England/-/emergency care hospitals

Health system characteristics: NR

Number of hospitals included in the analysis (intervention/control): 24/137

Characteristics of hospitals (before, intervention/control):

- Number of beds: NR
- Number of patients: NR
- Number of wards: NR
- Departments: NR
- Owner: NR
- Teaching status: NR
- Location:

Number of patients included in the analysis (intervention/control): 230,988/1,260,179

Characteristics of patients (intervention/control):

- Acute myocardial
 - * age (≥ 75 years): 43.1%/44%
 - * gender (female): 38.3%/36.7%
 - * casemix: NR
- Heart failure
 - * age (≥ 75 years): 61.4%/67.1%
 - * gender (female): 47%/48.7%
 - * casemix: NR
- Pneumonia
 - * age (≥ 75 years): 50.4%/52.2%
 - * gender (female): 49.6%/47.8%
 - * casemix: NR
- Indications:
 - * acute myocardial: 31.5%/31.0%
 - * heart failure: 24.1%/24.6%
 - * pneumonia: 44.4%/44.4%

Existing/other quality programs: intervention hospitals take part in a quality program (existing), investments in additional staff in the specialties covered if necessary independent from P4P bonus (during intervention)

Other relevant context information: none

Interventions

Period 1 (18 months): The first year was run as a pure tournament, with hospitals scoring in the top quartile on the quality metrics linked to incentives receiving a 4% bonus payment and those in the second quartile receiving a bonus of 2%. For the next 6 months, financial incentives were awarded on the basis of 3 criteria. Providers whose performance in this period was ranked above the median score from the first year were awarded an “attainment” bonus. Those earning this attainment bonus were then eligible for 2 further payments, which were awarded to hospitals in the top quartile for improved performance and those in the top 2 quartiles for absolute performance. Bonuses of USD 5 million (GBP 3.2

Kristensen 2014 (Continued)

million) were paid to hospitals in the northwest region for the first year and bonuses of USD 2.5 million (GBP 1.6 million) were paid for the next 6 months.

Period 2 (24 months): a fixed proportion of the hospital's expected income was withheld and paid out only if required performance thresholds were reached. The performance indicators remained the same, and required levels of achievement were based on the quality scores that had been achieved by each hospital in the first year of the Advancing Quality program. The total potential losses for hospitals were USD 5 million each year if all hospitals failed to meet all the targets for the 5 conditions under evaluation.

- Based on process indicators (e.g. discharge instructions provided)

Control: no P4P

Outcomes	In-hospital mortality during the first 30 days after admission	
Notes	Funding/conflict of interest: Supported by the National Institute for Health Research and the Danish Council for Independent Research, Social Sciences. 2 of the authors have a conflict of interest (grants from the funding organization). Information for subgroup analysis: no Other comments: -	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention cannot be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Mortality
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Low risk	No evidence for other risk of bias
Baseline outcomes similar All outcomes	High risk	Outcome measurements differ
Free of contamination	Low risk	It is unlikely that the control received the intervention
Baseline characteristics similar	Low risk	Adjusted for in the analysis

Kruse 2012

Methods	<p>Study design (CBA, ITS): CBA</p> <p>Allocation linked to study or natural experiment: linked to study</p> <p>Selection of hospitals (intervention/control):</p> <ul style="list-style-type: none"> Intervention: hospitals were invited (participation was voluntary) Control: hospitals not subject to P4P. Propensity-score matching to select a group of non-Premier comparison hospitals that were similar to the demonstration hospitals with respect to observed characteristics <p>Unit of allocation (region, hospital, country): hospitals</p> <p>Nature of desired change: initiation of P4P</p> <p>Data source: MedPAR files to identify AMI hospitalizations.</p> <p>Health care utilization over the 1 year after each AMI admission was tracked using the Standard Analytic Files containing claims for institutional outpatient providers, home health agencies, individual providers, and durable medical equipment and the 100% MedPAR file containing claims for all hospitalizations, skilled nursing facilities, and inpatient rehabilitation. These data were supplemented with the 100% Denominator File to identify HMO enrolment, patient date of birth, demographics, and death.</p> <p>Data were supplemented with Medicare claims data with hospital-level data from the annual Medicare Cost Reports, a variety of sources for hospital characteristics, and publicly available data on hospital performance from the CMS Website Hospital</p> <p>Compare for calculation of P4P bonus payments.</p> <p>Unit of analyses: hospitals</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): 2/0/2 (years)</p> <p>Statistical analyses:</p> <ul style="list-style-type: none"> Method: Logistic regression was used to estimate the propensity score. Difference in difference using ordinary least squares regression. Adjustment factors: <p>Propensity score: number of beds, ownership status, teaching status, accreditation by the Joint Commission, registered nurse- and licensed practical nurse-to-bed ratios, percentage of Medicare admissions, urban or rural location, the percentage of a hospital's patient days that are attributable to low-income patients, and level of market competition using the Herfindahl-Hirschman Index. Level of and quality as well as the change in these 2 factors over the 4 years prior to the initiation of P4P.</p> <p>Analysis: patient demographics (age, gender) and comorbidities, area-level characteristics such as market competition (Herfindahl-Hirschman Index)</p>
Participants	<p>Country/region/setting: USA/NR/hospitals</p> <p>Health system characteristics: Medicare and Medicaid patients</p> <p>Number of hospitals included in the analysis (intervention/control): 260/760</p> <p>Characteristics of hospitals (intervention/control):</p> <ul style="list-style-type: none"> Number of beds: 260/234 Number of patients: NR Number of wards: NR Departments: NR

Kruse 2012 (Continued)

- Owner:
 - * for profit: 1.9% /2.2%
 - * not-for-profit: 86.5% /85.0%
- Teaching status (medical school affiliation): 42.7%
- Location (urban): 36.5%/44.0%

Number of patients included in the analysis (before/after or intervention/control): NR

Characteristics of patients (intervention/control):

- Age: NR
- Gender: NR
- Casemix (mean): 1.42/1.39
- Indications: primary diagnosis of AMI

Existing/other quality programs:

Other relevant context information: none

Interventions	<p>The demonstration tracks hospitals' performance on measures related to the treatment of 5 conditions. These measures are combined into condition-specific composite scores, which are used to determine bonus payments. Financial bonuses were distributed as add-ons to diagnosis-related group (DRG) base payments for each targeted clinical condition in hospitals with performance in the top 20%. CMS paid participating hospitals more than USD 17 million in rewards in the first 2 years of the demonstration.</p> <p>Control: no P4P</p>
Outcomes	Hospital costs (per admission, hospital costs were calculated by converting MedPAR charges to costs using Medicare Cost Reports, cost-center-specific charge-to-cost ratios and summing for all hospitalizations)
Notes	Funding/conflict of interest: no disclosure

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention cannot be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Hospital cost
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting

Pay for performance for hospitals (Review)

Kruse 2012 (Continued)

Other bias	Low risk	No evidence for other risk of bias
Baseline outcomes similar All outcomes	Low risk	Matched sample
Free of contamination	Low risk	It is unlikely that the control received the intervention
Baseline characteristics similar	Low risk	Matched sample

Kwong 2017

Methods	<p>Study design (CBA, ITS): CBA</p> <p>Allocation linked to study or natural experiment: natural experiment</p> <p>Selection of hospitals (intervention/control): Medicare/non-Medicare population</p> <p>Unit of allocation (region, hospital, country): individual</p> <p>Nature of desired change: initiation</p> <p>Data source: data from the Healthcare Cost and Utilization Project (HCUP) National Inpatient Sample (NIS)</p> <p>Unit of analyses: individual</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): 18/-/10 (quarters)</p> <p>Statistical analyses:</p> <ul style="list-style-type: none"> • Method: Difference-in-difference, Poisson regression model • Adjustment factors: age, gender, race, elective admission, comorbidities, household income, teaching hospital, and urban or rural location
Participants	<p>Country/region/setting: USA/whole country/hospitals</p> <p>Health system characteristics: -</p> <p>Number of hospitals included in the analysis (before/after or intervention/control): NR</p> <p>Characteristics of hospitals (intervention/control):</p> <ul style="list-style-type: none"> • Number of beds: NR • Number of patients: NR • Number of wards: NR • Departments: NR • Owner: NR • Teaching status/location: <ul style="list-style-type: none"> * rural: 6.0/4.4 * urban non-teaching: 42.9/41.6 * urban teaching or missing: 51.1/54.0 <p>Number of patients included in the analysis (total): 1,753,854 discharges</p> <p>Characteristics of patients (intervention/control): spine fusion, shoulder and elbow repair, spinal re-fusion</p> <ul style="list-style-type: none"> • Age (mean): 70.6/64.1

Kwong 2017 (Continued)

- Gender (female): 59.0%/51.9%
- Casemix: NR
- Indications: NR

Existing/other quality programs: NR

Other relevant context information: none

Interventions	Nonpayment for surgical site infections: hospitals could no longer use a higher-level Medical Severity Diagnosis-Related Group (MS-DRG) denoting a complication that would result in higher reimbursements if the complication occurred after admission Control: no P4P
Outcomes	Surgical site infections
Notes	Funding/conflict of interest: Funding for this study was provided by the funders of Stanford MedScholars program. Dr Bhattacharya was partially funded by the National Institute on Aging; all authors have no conflicts of interest to disclose Information for subgroup analysis: no Other comments:-

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention cannot be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Surgical site infections
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not sufficiently reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Unclear risk	Analyses outcome same as outcome linked to payment
Baseline outcomes similar All outcomes	High risk	Difference in baseline outcomes
Free of contamination	High risk	Allocation on individual level. Medicare and non-Medicare patients can be treated in 1 hospital. Both might receive infection prevention measures

Kwong 2017 (Continued)

Baseline characteristics similar	High risk	Difference in baseline characteristics
----------------------------------	-----------	--

Lalloué 2017

Methods	<p>Study design (CBA, ITS): CBA</p> <p>Allocation linked to study or natural experiment: allocation linked to study</p> <p>Selection of hospitals (intervention/control): Volunteer hospitals randomly selected after stratification by type and administrative regions or chosen directly by the hospital federations/ volunteer hospitals not selected.</p> <p>Hospitals with undocumented quality indicators or hospitals that were only conditionally accredited during the course of the pilot study were subsequently excluded from the sample</p> <p>Unit of allocation (region, hospital, country): hospitals</p> <p>Nature of desired change: initiation</p> <p>Data source: quality indicator reporting, medical and administrative databases</p> <p>Unit of analyses: hospitals</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): NR</p> <p>Statistical analyses:</p> <ul style="list-style-type: none"> • Method: OLS regression • Adjustment factors: Hospital characteristics
Participants	<p>Country/region/setting: France/whole country/acute care hospitals</p> <p>Health system characteristics: NR</p> <p>Number of hospitals included in the analysis (intervention/control): 185/192</p> <p>Characteristics of hospitals (before [whole population], intervention/control): NR</p> <ul style="list-style-type: none"> • Number of beds: • Number of patients: • Number of wards: • Departments: • Owner: • Teaching status: • Location: <p>Number of patients included in the analysis (before/after or intervention/control): NR</p> <p>Characteristics of patients (before [whole population], intervention/control): NR</p> <ul style="list-style-type: none"> • Age: • Gender: • Casemix: • Indications: <p>Existing/other quality programs: Public reporting</p>

Lalloué 2017 (Continued)

Other relevant context information: none

Interventions	<p>From 9 process quality indicators (QIs), an aggregated score was constructed as the weighted average, taking into account both achievement and improvement. Hospitals with scores above the median received a financial reward based on their ranking and budget. Hospitals receive from 0.3% to 0.5% of their annual budget with minimum and maximum payments of USD 56,000.</p> <p>Control: no P4P</p>
Outcomes	<p>Quality score (process quality):</p> <p>“IFAQ score” (calculated out of the following 9 available process quality indicators).</p> <ul style="list-style-type: none"> • Traceability of pain assessment • Quality and content of the medical record • Quality and content of the anaesthesia records • Quality and content of multidisciplinary meetings in oncology • Time elapsed before sending discharge letters • Screening for nutritional disorders • Focus priority topics standards • Composite index for evaluation of activities against nosocomial infections • Medical record digitization
Notes	<p>Funding/conflict of interest: The work was supported by the French Ministry of Health and the National Authority for Health; conflict of interest not reported</p> <p>Information for subgroup analysis: no</p> <p>Other comments: -</p>

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention cannot be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Quality score
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Unclear risk	Analyzed outcome same as outcome linked to payment
Baseline outcomes similar	Unclear risk	Difference in baseline outcomes

Pay for performance for hospitals (Review)

Lalloué 2017 (Continued)

All outcomes

Free of contamination	Low risk	Allocation on hospital level
Baseline characteristics similar	Unclear risk	Not reported

Lee 2012

Methods	<p>Study design (CBA, ITS): ITS</p> <p>Allocation linked to study or natural experiment: natural experiment</p> <p>Selection of hospitals (intervention/control): subject to the Centers for Medicare and Medicaid Services inpatient prospective payment system rule and that reported data to the National Healthcare Safety Network before October 2008; all adult intensive care units or step-down units reported data on at least 1 of the 2 healthcare-associated infections of interest: central catheter associated bloodstream infections, catheter-associated urinary tract infections</p> <p>Unit of allocation (region, hospital, country): NA</p> <p>Nature of desired change: initiation of P4P</p> <p>Data source: 2009 American Hospital Association annual survey</p> <p>Unit of analyses: hospital units</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): 11/1/9 (quarter)</p> <p>Statistical analyses:</p> <ul style="list-style-type: none"> • Method: negative binomial mixed-effects model • Adjustment factors: time, interaction between time and intervention
Participants	<p>Country/region/setting: USA/41 states/ nonfederal acute care hospitals</p> <p>Health system characteristics: Medicare and Medicaid patients</p> <p>Number of hospitals included in the analysis (before/after): 398/398</p> <p>Characteristics of hospitals (before):</p> <ul style="list-style-type: none"> • Number of beds: <ul style="list-style-type: none"> * < 100: 58 (15%) * 100 to < 400: 235 (59%) * ≥ 400: 105 (26%) • Number of patients: <ul style="list-style-type: none"> * Medicare admissions (median, range): 45% (39% to 51%) * Medicaid admissions (median, range): 17% (13% to 22%) • Number of wards: NR • Departments: NR • Owner: <ul style="list-style-type: none"> * public: 39 (10%) * for-profit: 47 (12%) * not-for-profit: 312 (78%)

Lee 2012 (Continued)

- Teaching status:
 - * major: 78 (20%)
 - * graduate: 89 (22%)
 - * limited: 23 (6%)
 - * nonteaching: 198 (50%)
 - * data missing: 10 (3%)
- Location:
 - * metropolitan: 337 (85%)
 - * micropolitan: 45 (11%)
 - * rural: 16 (4%)

Number of patients included in the analysis (before/after or intervention/control): NR

Characteristics of patients (before):

- Age: NR
- Gender: NR
- Casemix: NR
- Indications:
 - * central catheter-associated bloodstream infection: 392 (98%)
 - * catheter-associated urinary tract infection: 245 (62%)

Existing/other quality programs: some states with mandatory reporting of infection rates

Other relevant context information: none

Interventions	Nonpayment (no reimbursement) for health care-acquired conditions. Control: No P4P
Outcomes	Healthcare-associated infection (measurement not reported, rate per 1000 device-days exposed)
Notes	Funding/conflict of interest: NR Information for subgroup analysis: size (admissions), monitoring (reporting) Other comments:

Risk of bias

Bias	Authors' judgement	Support for judgement
Intervention independent (ITS)	Unclear risk	Not enough information provided
Shape of effect pre-specified (ITS)	Low risk	Rationale for shape is given
Unlikely to affect data collection (ITS) All outcomes	High risk	Analysed outcome same as outcome linked to payment
Knowledge of the allocated interventions, Blinding (ITS, objective outcomes)	Low risk	Healthcare-associated infection
Incomplete outcome data addressed (ITS)	Low risk	Only 3% missing data. Unlikely to bias results

Lee 2012 (Continued)

All outcomes

Free of selective reporting (ITS)	Low risk	No indication for selective reporting
Free of other bias (ITS)	Low risk	No evidence for other bias

McGarry 2016

Methods	<p>Study design (CBA, ITS): CBA</p> <p>Allocation linked to study or natural experiment: natural experiment</p> <p>Selection of hospitals (intervention/control):</p> <ul style="list-style-type: none"> Intervention: eligible for Hospital Readmission Reduction Program. Only hospitals eligible for penalties for all 3 target conditions Control: unclear <p>Unit of allocation (region, hospital, country): hospital</p> <p>Nature of desired change: implementation</p> <p>Data source: Study data comes from the NYS Statewide Planning and Research Cooperative System hospital claims database spanning years 2008 to 2013. Information on hospital characteristics and whether a facility was eligible for Hospital Readmission Reduction Program were obtained from the 2012 Medicare Impact File and the 2012 Hospital Compare data archive. ZIP code-level socioeconomic information was obtained from the 2012 American Community Survey.</p> <p>Unit of analyses: hospitals</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): NR</p> <p>Statistical analyses:</p> <ul style="list-style-type: none"> Method: Difference-in-difference analysis. Logistic regression with hospital-level random effects. Adjustment factors: demographics and a variety of clinical factors related to the index admission, including length of stay, service category, whether the admission originated in an ED, a categorical severity of illness measure, discharge location, discharge occurs on the weekend, comorbidity, socioeconomic status, hospital size (bed count), teaching hospital, hospital ownership status and hospital location (urban/rural). In addition time trend.
Participants	<p>Country/region/setting: USA/whole country/hospitals</p> <p>Health system characteristics: Medicare</p> <p>Number of hospitals included in the analysis (total): 141</p> <p>Characteristics of hospitals (total):</p> <ul style="list-style-type: none"> Number of beds (bed size, mean, SD): 350.4 (334) Number of patients: NR Number of wards: NR Departments: NR Owner (non-for-profit): 88.7% Teaching status (teaching): 61% Location (rural): 12.1% <p>Number of patients included in the analysis (total): 229,358 discharges</p>

McGarry 2016 (Continued)

Characteristics of patients (total): > 65 years

- Age (mean, SD): 81.1 (8.5)
- Gender: 55.6 %
- Casemix: NR
- Indications: acute myocardial infarction, heart failure, and pneumonia

Existing/other quality programs: public reporting for readmissions

Other relevant context information: none

Interventions	<p>The Hospital Readmission Reduction Program penalizes hospitals with adjusted readmission rates that are higher than the national average through a reduction in base IPPS Medicare payments. Adjusted readmission rates account for patient demographics and severity of illness. The data for calculating hospital-level readmission rates come from Medicare FFS claims over a 3-year period. For example, for the initial penalties administered in October 2012, readmission rates were determined using data from June 2008 to July 2011. Penalties were initially capped at 1% of base inpatient claims with scheduled increases up to 3% by fiscal year 2015. Similarly, readmission rates were calculated for 3 conditions — acute myocardial infarction, heart failure, and pneumonia — at program outset; target conditions were expanded in fiscal year 2015 to include hip and knee replacements, as well as chronic obstructive pulmonary disease.</p> <p>Control: Hospital with lower percentage of (Medicare) patients that might be affected by penalties.</p>
Outcomes	Readmissions within 30 days, ED visits within 30 days
Notes	<p>Funding and conflict of interest: Supported by the New York State Department of Health. The authors declare no conflict of interest.</p> <p>Information for subgroup analysis: -</p> <p>Other comments: -</p>

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, Intervention cannot be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Readmission objective outcome
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Handling missing outcome data not reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting

McGarry 2016 (Continued)

Other bias	High risk	Analysed outcome same as outcome linked to payment
Baseline outcomes similar All outcomes	Unclear risk	Baseline outcome measures per group not reported
Free of contamination	Low risk	Allocation on hospital level
Baseline characteristics similar	Unclear risk	Baseline characteristics not reported per group

Mellor 2017

Methods	<p>Study design (CBA, ITS): CBA</p> <p>Allocation linked to study or natural experiment: natural experiment</p> <p>Selection of hospitals (intervention/control): Medicare patients/private insured patients</p> <p>Unit of allocation (region, hospital, country): individual</p> <p>Nature of desired change: initiation</p> <p>Data source: Virginia Health Information, Hospital Compare data</p> <p>Unit of analyses: individual</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): 8/excluded/10 (quarter)</p> <p>Statistical analyses:</p> <ul style="list-style-type: none"> - Method: difference in difference, hospital random effects, logistic or linear regression - Adjustment factors: race/ethnicity, for each age, and for female patients; plus, as proxies for patient health, the numbers of chronic conditions, comorbid conditions, and procedures performed, a full set of indicator variables for the patient's principal diagnosis, unemployment rate, residents in poverty, median household income, Medicare share, time period
Participants	<p>Country/region/setting: USA/Virginia/hospitals</p> <p>Health system characteristics: Medicare</p> <p>Number of hospitals included in the analysis (before/after or intervention/control): NR</p> <p>Characteristics of hospitals (before [whole population], intervention/control): NR</p> <ul style="list-style-type: none"> • Number of beds: • Number of patients: • Number of wards: • Departments: • Owner: • Teaching status: • Location: <p>Number of patients included in the analysis (total): 36,193 (myocardial infraction); 69,713 (heart failure); 50,220 (pneumonia)</p> <p>Characteristics of patients (before [whole population], intervention/control): NR</p> <ul style="list-style-type: none"> • Age: NR

Mellor 2017 (Continued)

- Gender: NR
- Casemix: NR
- Indications: myocardial infarction, heart failure, pneumonia

Existing/other quality programs: not reported

Other relevant context information: none

Interventions	<p>Reduced payments to hospitals with excess 30-day readmissions for Medicare patients treated for acute myocardial infarction, heart failure, and pneumonia. 1% of total payments and 64% of hospitals were penalized in the first year. Average reduction among penalized hospitals was 0.42%. 3 years later the maximum penalty increased to 3% and applicable conditions also included chronic obstructive pulmonary disease and total hip and knee replacements.</p> <p>Penalties are based on 'excess readmissions' (relative to a national mean). Penalties are based on a hospital's 3-year average excess readmission rate</p> <p>Control:</p> <p>No P4P</p>
Outcomes	Readmission (31 to 45 days)
Notes	<p>Funding/conflict of interest: The research was supported by the Schroeder Center for Health Policy at the College of William and Mary; the authors have no conflict of interest</p> <p>Information for subgroup analysis: intensity</p> <p>Other comments: no</p>

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention cannot be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Readmission
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Unclear risk	Analyses outcome same as outcome linked to payment
Baseline outcomes similar All outcomes	Unclear risk	Not reported

Pay for performance for hospitals (Review)

Mellor 2017 (Continued)

Free of contamination	High risk	Allocation on individual level. Medicare and private insured patients can be treated in 1 hospital. Both might receive infection prevention measures
Baseline characteristics similar	Unclear risk	Not reported

Morgan 2012

Methods	<p>Study design (CBA, ITS): ITS</p> <p>Allocation linked to study or natural experiment: natural experiment</p> <p>Selection of hospitals (intervention/control): Member hospitals of the Society for Healthcare Epidemiology of America Research Network, a consortium of > 200 hospitals that has successfully conducted multicenter research projects in healthcare epidemiology</p> <p>Unit of allocation (region, hospital, country): NA</p> <p>Nature of desired change: implementation</p> <p>Data source: hospitals provided daily retrospective data for the study period</p> <p>Unit of analyses: hospitals</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): unclear</p> <p>Statistical analyses:</p> <ul style="list-style-type: none"> - Method: Segmented regression. Poisson mixed-effects models to account for within-hospital correlation - Adjustment factors: NR
Participants	<p>Country/region/setting: USA/22 states/tertiary care hospitals</p> <p>Health system characteristics:</p> <p>Number of hospitals included in the analysis (total): 10</p> <p>Characteristics of hospitals (before):</p> <ul style="list-style-type: none"> • Number of beds (range): 120 to 1000 • Number of patients: NR • Number of wards: NR • Departments: NR • Owner: NR • Teaching status: NR • Location: NR <p>Number of patients included in the analysis (total): 2 362 742 admissions</p> <p>Characteristics of patients (before [whole population], intervention/control):</p> <ul style="list-style-type: none"> • Age: NR • Gender: NR • Casemix: NR • Indications: NR <p>Existing/other quality programs: NR</p>

Morgan 2012 (Continued)

Other relevant context information: none

Interventions	Hospitals are not reimbursed for catheter-associated urinary tract infection. Control: no penalties
Outcomes	Antimicrobial use (in hospital)
Notes	Funding and conflict of interest: The work was supported by the Society for Hospital Epidemiology of America. 2 authors have received an unrestricted research grant from Merck and have served on a speakers' bureau for Merck. 2 authors have received payment for contributions to UpToDate Online. All other authors report no potential conflicts. Information for subgroup analysis: - Other comments: -

Risk of bias

Bias	Authors' judgement	Support for judgement
Intervention independent (ITS)	Unclear risk	No information
Shape of effect pre-specified (ITS)	Low risk	Shape pre-specified
Unlikely to affect data collection (ITS) All outcomes	High risk	Analysed outcome same as outcome linked to payment
Knowledge of the allocated interventions, Blinding (ITS, objective outcomes)	Low risk	Antimicrobial prescribing objective outcome. Outcome was validated in the study
Incomplete outcome data addressed (ITS) All outcomes	High risk	Only 10 of 35 hospitals provided requested antimicrobial prescribing data
Free of selective reporting (ITS)	Unclear risk	Focus on statistically significant results
Free of other bias (ITS)	Low risk	No evidence for other bias

Padula 2015

Methods	Study design (CBA, ITS): ITS Allocation linked to study or natural experiment: natural experiment Selection of hospitals (intervention/control): Members of the University HealthSystem Consortium Unit of allocation (region, hospital, country): NA Nature of desired change: initiation of P4P Data source: University HealthSystem Consortium's Clinical Data Base and Resource Manager administrative discharge data.
---------	---

Padula 2015 (Continued)

Unit of analyses: patients

Number of measurements (before, transition, after, unit [e.g. years]): 3/-/15 (quarter)

Statistical analyses:

- Method: ARIMA (reanalyzed)
- Adjustment factors: none

Participants	<p>Country/region/setting: USA/-/academic centers</p> <p>Health system characteristics: -</p> <p>Number of hospitals included in the analysis (before/after): 170/184 to 210 (depending on year)</p> <p>Characteristics of hospitals (before [whole population], intervention/control):</p> <ul style="list-style-type: none"> • Number of beds: NR • Number of patients: NR • Number of wards: NR • Departments: NR • Owner: NR • Teaching status: NR • Location: NR <p>Number of patients included in the analysis (before/after): 829,316/3,251,056 (discharges)</p> <p>Characteristics of patients (before): at least 5 days' hospitalization</p> <ul style="list-style-type: none"> • Age: <ul style="list-style-type: none"> * 18 to 30: 9.39% * 31 to 50: 24.60% * 51 to 63: 27.55% * > 63: 38.58% • Gender (female): 49.71% • Casemix (mean): 2.28 • Indications: NR <p>Existing/other quality programs: NR</p> <p>Other relevant context information: none</p>
Interventions	<p>Non-payment (no reimbursement) for hospital-acquired pressure ulcers</p> <p>Control: no nonpayment</p>
Outcomes	<p>Hospital-acquired pressure ulcers (measured according to Agency for Healthcare Research and Quality Patient Safety Indicator)</p>
Notes	<p>Funding/conflict of interest: NR</p> <p>Information for subgroup analysis: -</p> <p>Other comments: -</p>
Risk of bias	
Bias	Authors' judgement Support for judgement

Padula 2015 (Continued)

Intervention independent (ITS)	Unclear risk	Not enough information provided
Shape of effect pre-specified (ITS)	Low risk	Rationale for shape is given
Unlikely to affect data collection (ITS) All outcomes	High risk	Analysed outcome same as outcome linked to payment
Knowledge of the allocated interventions, Blinding (ITS, subjective outcomes)	Unclear risk	Hospital-acquired pressure ulcers, measurement not specified
Incomplete outcome data addressed (ITS) All outcomes	Unclear risk	Not reported
Free of selective reporting (ITS)	Low risk	No indication for selective reporting
Free of other bias (ITS)	Low risk	No evidence for other bias

Ryan 2009

Methods	<p>Study design (CBA, ITS): CBA</p> <p>Allocation linked to study or natural experiment: linked to study</p> <p>Selection of hospitals (intervention/control): Hospitals' eligibility to participate was based on their subscription to Premier's Perspective database, a database used for benchmarking and quality improvement activities</p> <ul style="list-style-type: none"> Intervention: voluntary participation Control: hospitals declined to participate <p>Unit of allocation (region, hospital, country): hospital</p> <p>Nature of desired change: initiation of P4P</p> <p>Data source: Medicare data from 2000 to 2006: inpatient claims, Denominator files, and Provider of Service files. Inpatient claims are used to identify the principal diagnoses for which beneficiaries are admitted, secondary diagnoses and type of admission for risk adjustment, cost data, and discharge status to exclude transfer patients. The Medicare Denominator File is used to add additional risk adjusters and to determine 30-day mortality. Data from the Medicare Provider of Service file are used to identify hospital structural characteristics</p> <p>Unit of analyses: hospitals</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): 3/-/3 (years)</p> <p>Statistical analyses:</p> <ul style="list-style-type: none"> Method: difference-indifference (multivariate linear regression) Adjustment factors: age, gender, race, 30 dummy variables for comorbidities, type of admission (emergency, urgent, elective), and season of admission
Participants	<p>Country/region/setting: USA/-/acute care hospitals</p>

Pay for performance for hospitals (Review)

Ryan 2009 (Continued)

Health system characteristics: Medicare and Medicaid patients

Number of hospitals included in the analysis (intervention/control): 256/116

Characteristics of hospitals (intervention/control): Only short-term, acute care hospitals are included in the analysis

- Number of beds:
 - * 1 to 99: 12.1/19.5
 - * 100 to 399: 57.8/61.0
 - * ≥400: 30.1/19.5
- Number of patients: NR
- Number of wards: NR
- Departments: NR
- Owner:
 - * not-for-profit: 85.2%/83.8%
 - * for-profit: 2.3%/4.3%
 - * government run: 12.5%/12.0%
- Teaching status (medical school affiliation): 43.4/28.0
- Location: NR

Number of patients included in the analysis (total): 6,713,928 patients, 11,232,452 admissions

Characteristics of patients (before [whole population], intervention/control):

- Age: NR
- Gender: NR
- Casemix: NR
- Indications: AMI, heart failure, CBAG, pneumonia

Existing/other quality programs: NR

Other relevant context information: none

Interventions	2% bonus on Medicare reimbursement rates to hospitals performing in the top decile of performance of a composite quality measure for each clinical condition incentivized and a 1% bonus for hospitals performing in the second highest decile. Penalties for very low performing hospitals were implemented in 2006 (year of last observation)
	Control: No P4P
Outcomes	Risk adjusted hospital cost (60 days, measurement not specified), mortality (30 days)
Notes	Funding/conflict of interest: training grant from Agency for Health Care Research and Quality, no disclosure Information for subgroup analysis: - Other comments: -

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA

Pay for performance for hospitals (Review)

Ryan 2009 (Continued)

Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention cannot be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Hospital costs
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Low risk	No evidence for other risk of bias
Baseline outcomes similar All outcomes	High risk	Outcome measurements differ
Free of contamination	Low risk	It is unlikely that the control received the intervention
Baseline characteristics similar	Low risk	Adjusted for in the analysis

Ryan 2010

Methods	<p>Study design (CBA, ITS): CBA</p> <p>Allocation linked to study or natural experiment: linked to study</p> <p>Selection of hospitals (intervention/control): Hospitals' eligibility to participate was based on their subscription to Premier's Perspective database, a database used for benchmarking and quality improvement activities.</p> <ul style="list-style-type: none"> Intervention: voluntary participation Control: NR <p>Unit of allocation (region, hospital, country): hospitals</p> <p>Nature of desired change: initiation of P4P</p> <p>Data source: Medicare data from 2000 to 2006 in this analysis: 100% inpatient claims, denominator files, and provider of service files. Inpatient claims are used to identify the principal diagnoses for which beneficiaries are admitted and secondary diagnoses and type of admission for risk adjustment. The Medicare denominator file is used to include additional risk adjusters and to determine beneficiary zip code of residence. Data from the Medicare provider of service file are used to identify hospital structural characteristics.</p> <p>Unit of analyses: patients</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): 8/-/8 (quarter)</p> <p>Statistical analyses:</p> <ul style="list-style-type: none"> Method: difference-in-difference (multiple linear regression)
---------	---

Ryan 2010 (Continued)

- Adjustment factors: age, gender, 30 dummy variables' comorbidities, type of admission (emergency, urgent, elective), and season of admission

Participants	<p>Country/region/setting: USA/-/hospitals</p> <p>Health system characteristics: NR</p> <p>Number of hospitals included in the analysis (total): 1063</p> <p>Characteristics of hospitals (before [whole population], intervention/control): Only short-term, acute care hospitals are included in the analysis</p> <ul style="list-style-type: none">• Number of beds: NR• Number of patients: NR• Number of wards: NR• departments: NR• Owner: NR• Teaching status: NR• Location: NR <p>Number of patients included in the analysis (before/after or intervention/control): NR</p> <p>Characteristics of patients (before [whole population], intervention/control):</p> <ul style="list-style-type: none">• Age: NR• Gender: NR• Casemix: NR• Indications: AMI <p>Existing/other quality programs: NR</p> <p>Other relevant context information: none</p>	
Interventions	<p>2% bonus on Medicare reimbursement rates to hospitals performing in the top decile of a composite quality measure for each incentivized condition and a 1% bonus for hospitals performing in the second decile. Penalties were administered to hospitals with exceptionally poor performance. Bonus payments were disbursed based on composite quality measures, consisting predominately of process measures but including some outcome measures, for each incentivized condition</p> <p>Control: no P4P</p>	
Outcomes	Difference in CBAG	
Notes	<p>Funding and conflict of interest: supported by the Jewish Healthcare Foundation under the grant “Achieving System-wide Quality Improvements”. No conflict of interest disclosure.</p> <p>Information for subgroup analysis: -</p> <p>Other comments: -</p>	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA

Ryan 2010 (Continued)

Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention cannot be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Difference in CBAG
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Low risk	No evidence for other risk of bias
Baseline outcomes similar All outcomes	High risk	Outcome measurements differ
Free of contamination	Low risk	It is unlikely that the control received the intervention
Baseline characteristics similar	Low risk	Adjusted for in analysis

Ryan 2011

Methods	<p>Study design (CBA, ITS): CBA</p> <p>Allocation linked to study or natural experiment: natural experiment</p> <p>Selection of hospitals (intervention/control): hospitals in the Massachusetts Medicaid program.</p> <p>Unit of allocation (region, hospital, country): region</p> <p>Nature of desired change: implementation</p> <p>Data source: data on all-payer hospital process of care performance from Medicare's Hospital Compare program and data on hospital characteristics from Hospital Compare and the 2005 American Hospital Association Annual Survey.</p> <p>Unit of analyses: hospitals</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): pneumonia 4/-/2; surgical site infection 5/-/1 (years)</p> <p>Statistical analyses:</p> <ul style="list-style-type: none"> • Method: difference-in-difference using linear regression • Adjustment factors: time trends, outcome measure completeness
Participants	<p>Country/region/setting: USA/-/acute care hospitals and critical access hospitals</p> <p>Health system characteristics: Medicare</p> <p>Number of hospitals included in the analysis (intervention/control): 62/3676</p>

Ryan 2011 (Continued)

Characteristics of hospitals (intervention/control): Acute care hospitals; Critical Access Hospitals (CAHs), small, rural hospitals were excluded

- Number of beds:
 - * 1 to 99: 23%/33%
 - * 100 to 399: 65%/60%
 - * > 400: 12%/7%
- Number of patients:
- Number of wards: NR
- Departments: NR
- Owner:
 - * government: 5%/20%
 - * for-profit: 2%/20%
 - * not for profit: 94%/61%
- Teaching status (teaching): 22%/8%
- Location (Urban): 88%/66%

Number of patients included in the analysis (total): 19,569 (pneumonia); 13,678 (surgical site infection)

Characteristics of patients (before [whole population], intervention/control):

- Age: NR
- Gender: NR
- Casemix: NR
- Indications: NR

Existing/other quality programs: public reporting

Other relevant context information: none

Interventions	<p>Incentives for pneumonia and surgical infection prevention. Hospitals were rewarded based on composite process quality measures calculated separately for each condition. Payments were calculated as the product of a hospital's quality score, its number of eligible opportunities, and a predetermined dollar amount, which varies across incentivized conditions. In 2008, up to USD 4.5 million in incentives were available statewide for pneumonia quality, with USD 2.6 million ultimately disbursed in payments to hospitals, averaging approximately USD 40,000 per hospital (25th percentile = USD 10,942; 50th percentile = USD 27,356; 75th percentile = USD 57,447)</p> <p>Control: no P4P</p>
Outcomes	<p>Process composite quality measure</p> <ul style="list-style-type: none"> • Pneumonia: oxygenation assessment, blood culture performed in emergency department before first antibiotic received in hospital, adult smoking cessation advice and counseling, initial antibiotic received within 6 hours of arrival, and appropriate antibiotic selection in immunocompetent patients • Surgical site infection: prophylactic antibiotic within 1 hour of surgical incision, appropriate antibiotic selection for surgical prophylaxis, and prophylactic antibiotic discontinued within 24 hours after surgery end-time
Notes	<p>Funding/conflict of interest: For Andrew Ryan, this work has been supported by a K01 career development award from Agency for Health Care Research and Quality</p> <p>Information for subgroup analysis: -</p> <p>Other comments:</p>
Risk of bias	

Ryan 2011 (Continued)

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention cannot be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Process composite quality measure
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Unclear risk	Analyses outcome same as outcome linked to payment
Baseline outcomes similar All outcomes	High risk	Outcome measurements differ
Free of contamination	Low risk	It is unlikely that the control received the intervention
Baseline characteristics similar	Unclear risk	Not all relevant characteristics are reported

Ryan 2012a

Methods

Study design (CBA, ITS): CBA

Allocation linked to study or natural experiment: linked to study

Selection of hospitals (intervention/control):

- Intervention: voluntary participation
- Control: matched sample of US hospitals not participating in the demonstration

Unit of allocation (region, hospital, country): hospitals

Nature of desired change: modification of P4P

Data source: hospital-level data on quality from Hospital Compare for discharges; data on hospital characteristics from the American Hospital Association Annual Survey; data on the receipt of incentive payments from the Premier website; and, to estimate incentive payments in phase 1, data on hospital revenues from the Medicare Provider Analysis and Review files

Unit of analyses: hospitals

Number of measurements (before, transition, after, unit [e.g. years]): 3-/3 (years)

Ryan 2012a (Continued)

Statistical analyses:

- Method: difference-in-differences (multiple linear regression)
- Adjustment factors:
- Matching criteria: baseline quality
- Adjustment factors: baseline quality

Participants

Country/region/setting: USA/-/acute care hospitals

Health system characteristics: Medicare and Medicaid patients

Number of hospitals included in the analysis (intervention/control): 250/250

Characteristics of hospitals (intervention/control):

- Number of beds:
 - * 1 to 99: 12.4%/10.0%
 - * 100 to 399: 72.0%/76.4%
 - * ≥ 400: 15.6%/13.6%
- Number of patients: NR
- Number of wards: NR
- Departments: NR
- Owner:
 - * not-for-profit: 87.6%/84.8%
 - * government run: 10.8%/3.2%
 - * for-profit: 1.6%/2.0%
- Teaching status (medical school affiliation): 38.8%/37.2%
- Location (urban): 78.8%/78.0%

Number of patients included in the analysis (before/after or intervention/control): NR

Characteristics of patients (intervention/control):

- Age: NR
- Gender: NR
- Casemix: NR
- Indications:
 - AMI: 89.9%/90.1%
 - Heart failure: 77.4%/77.7%
 - Pneumonia: 78.6%/78.7%

Existing/other quality programs: public reporting

Other relevant context information: none

Interventions

Phase 1 (before): 2% bonus on its reimbursement rates to hospitals performing in the top tenth of demonstration hospitals on a composite quality measure for each clinical diagnosis and procedure incentivized in the demonstration; and a 1% bonus for hospitals performing in the second-highest decile. Penalties were imposed for hospitals performing below the twentieth percentile of hospitals 2 years prior to the current year.

Phase 2 (after): Hospitals were eligible to receive 3 types of rewards. First was an attainment award, given to hospitals whose composite scores in the current year exceeded the median of demonstration hospitals 2 years prior to the current year. Second was a top performer award, given to hospitals that scored in the top 20% of demonstration hospitals in the current year. Third was an improvement award, given to hospitals with scores above the median of demonstration hospitals in the current year that ranked in the top 20% of demonstration hospitals for quality improvement. Hospitals could receive both top performer and attainment awards or both improvement and attainment awards. However, they could not receive both top performer and improvement awards.

Ryan 2012a (Continued)

The amount of incentive payments increased from an average of USD 8.2 million per year in phase 1 to USD 12 million per year in phase 2. Of the phase 2 bonuses, 60% was allocated to top performer and improvement awards and 40% to attainment awards. The incentivized quality measures remained very similar across the 2 phases.

Control: no P4P

Outcomes	Compoents of the composite process quality score (sum of successfully achieved processes divid- ed by the number of patients eligible to receive these processes) <ul style="list-style-type: none">• AMI<ul style="list-style-type: none">* aspirin at arrival* aspirin prescribed at discharge* angiotensin-converting enzyme inhibitors (ACEI) <i>or</i> angiotensin receptor blocker (ARB) for Left Ven- tricular Systolic Dysfunction(LVSD)* adult smoking cessation advice/counseling* beta-blocker prescribed at discharge* beta-blocker at arrival* fibrinolytic received within 30 minutes of hospital arrival* primary PCI received within 90 minutes of hospital arrival• Heart failure:<ul style="list-style-type: none">* evaluation of LVS function* ACEI or ARB for LVSD* discharge instructions* adult smoking cessation advice/counseling• Pneumonia:<ul style="list-style-type: none">* oxygenation assessment* initial antibiotic selection in immunocompetent patients* blood cultures performed in the ED prior to initial antibiotic received in hospital* influenza vaccination* pneumococcal vaccination* initial antibiotic received within 6 hours of hospital arrival* adult smoking cessation advice/counseling	
Notes	Funding/conflict of interest: supported by Agency for Healthcare Research and Quality Information for subgroup analysis: Other comments: -	
<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement
Random sequence genera- tion (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (perfor- mance bias) All outcomes	High risk	CBA, intervention cannot be blinded
Blinding of outcome as- sessment (detection bias)	Low risk	Composite process quality

Ryan 2012a (Continued)

Objective outcomes

Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Low risk	No evidence of other risk of bias
Baseline outcomes similar All outcomes	Low risk	Matched samples
Free of contamination	Low risk	It is unlikely that the control received the intervention
Baseline characteristics similar	Low risk	Matched samples

Ryan 2015

Methods	<p>Study design (CBA, ITS): CBA</p> <p>Allocation linked to study or natural experiment: natural experiment</p> <p>Selection of hospitals (intervention/control): all acute care hospitals/all critical access hospitals and hospitals in another region</p> <p>Unit of allocation (region, hospital, country): hospital type and region</p> <p>Nature of desired change: initiation</p> <p>Data source: Quality of care to Hospital Compare, Medicare's public quality reporting initiative</p> <p>Unit of analyses: hospitals</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): up to 20/NA/3 (quarter)</p> <p>Statistical analyses:</p> <ul style="list-style-type: none"> • Method: matched control, difference-in-difference, linear regression • Adjustment factors: outcome values at baseline
Participants	<p>Country/region/setting: USA/whole country/acute care hospitals and critical access hospitals</p> <p>Health system characteristics: Medicare</p> <p>Number of hospitals included in the analysis (intervention/control): 2801/240 (process outcomes); 2779/284 (patient experience outcomes)</p> <p>Characteristics of hospitals (intervention/control):</p> <ul style="list-style-type: none"> • Number of beds: 231/82 (process outcomes); 238/130 (patient experience outcomes) • Number of patients: NR • Number of wards: NR • Departments: NR

Ryan 2015 (Continued)

- Owner:
 - * government: 17.0/22.9 (process outcomes); 17.7/16.7 (patient experience outcomes)
 - * for profit: 18.2/5.7 (process outcomes); 17.5/7.0 (patient experience outcomes)
 - * not for profit: 64.7/71.4 (process outcomes); 64.8/76.3 (patient experience outcomes)
- Teaching status: 32.9%/16.1% (process outcomes); 33.2%/21.6% (patient experience outcomes)
- Location: NR

Number of patients included in the analysis (total): 18,246 (process outcomes); 12,252 (patient experience outcomes)

Characteristics of patients (before [whole population], intervention/control):

- Age: NR
- Gender: NR
- Casemix: NR
- Indications: all

Existing/other quality programs: public reporting

Other relevant context information: none

Interventions	<p>Payment adjustments on 12 clinical process and 8 patient experience measures. Budget neutral, re-distributing hospital payment “withholds” from “losing” to “winning” hospitals. These withholds are equal to 1% of hospital payments from diagnosis related groups (DRGs) in the initial implementation period. Incentive payments based on a unique approach that incorporates both quality attainment and quality improvement, incentivizing hospitals for incremental improvements and foregoing the all-or-nothing threshold design of other programs.</p> <p>Incentivized Clinical Process of Care Measures: acute myocardial infarction; fibrinolytic therapy; primary percutaneous coronary intervention; heart failure; discharge instructions; pneumonia; blood cultures performed in the emergency department; initial antibiotic selection; surgical care improvement; prophylactic antibiotic received; prophylactic antibiotic selection; prophylactic antibiotics discontinued; cardiac patients with controlled 6AM postoperative serum glucose; venous thromboembolism prophylaxis ordered; appropriate venous thromboembolism prophylaxis 24 hours before and after surgery; appropriate continuation of beta blocker. Incentivized patient experience measures: communication with nurses; communication with doctors; responsiveness of hospital staff; pain management; communication about medicines; cleanliness and quietness of hospital environment; discharge information; overall rating of hospital.</p> <p>Control:</p> <p>No P4P</p>
Outcomes	<p>Clinical process performance (12 measures that were incentivized in the first year), percentage of opportunities for providing recommended care that were actually provided; patient experience performance (measures that were incentivized in the initial performance period), percentage of patients reporting “always” to each of the questions (e.g. patients who reported that their doctors “always” communicated well).</p>
Notes	<p>Funding/conflict of interest: Wood Johnson Foundation under the Changes in Health Care Financing and Organization program</p> <p>Authors report personal funding by policy organizations and company that provides software and analytic services for assessing hospital quality and efficiency</p> <p>Information for subgroup analysis: low initial performance</p> <p>Other comments: none</p>

Risk of bias

Ryan 2015 (Continued)

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention can not be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Process outcomes
Blinding of outcome assessment (detection bias, subjective outcomes)	Unclear risk	Patient experience: blinding not specified
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not sufficiently reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Unclear risk	Analyses outcome same as outcome linked to payment
Baseline outcomes similar All outcomes	Low risk	Propensity score matching
Free of contamination	Low risk	Allocation on hospital level
Baseline characteristics similar	Low risk	Propensity score matching

Ryan 2017

Methods	<p>Study design (CBA, ITS): CBA</p> <p>Allocation linked to study or natural experiment: natural experiment</p> <p>Selection of hospitals (intervention/control): all acute care hospitals/all critical access hospitals</p> <p>Unit of allocation (region, hospital, country): type of hospital</p> <p>Nature of desired change: initiation</p> <p>Data source: data from Hospital Compare</p> <p>Unit of analyses: hospitals</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): 4/-/3 (years)</p> <p>Statistical analyses:</p>
---------	--

Ryan 2017 (Continued)

- Method: difference-in-difference, linear fixed-effects model
- Adjustment factors: outcome baseline values

Participants	<p>Country/region/setting: USA/whole country/hospitals</p> <p>Health system characteristics: Medicare</p> <p>Number of hospitals included in the analysis (intervention/control): 2164/153 (clinical process outcome); 1507/237 (patient experience); 1364/31 (mortality myocardial infarction); 2383/419 (mortality heart failure); 2615/617 (mortality pneumonia)</p> <p>Characteristics of hospitals (intervention/control): short-term acute care hospitals/critical access hospitals</p> <ul style="list-style-type: none"> • Number of beds: (mean, SD): 223(187)/24(2) (clinical process outcome); 177(170)/24(2) (patient experience); 242(182)/25(0) (mortality myocardial infarction); 203(183)/24(3) (mortality heart failure); 208(187)/24(4) (mortality pneumonia) • Number of patients: NR • Number of wards: NR • Departments: NE • Owner: NR • Teaching status (teaching): 37/1; 28/2; 39/0; 32;1 33;1 • Location: NR <p>Number of patients included in the analysis (before/after or intervention/control): NR</p> <p>Characteristics of patients (before [whole population], intervention/control): mortality outcomes: patients who were admitted to the hospital for acute myocardial infarction, heart failure, and pneumonia</p> <ul style="list-style-type: none"> • Age: NR • Gender: NR • Casemix: NR • Indications: <p>Existing/other quality programs: public reporting</p> <p>Other relevant context information: none</p>
Interventions	<p>Starting with clinical-process and patient-experience measures, the program expanded to include patient outcome measures and spending measures. The size of the program incentives has also increased gradually from 1% of diagnosis-related group revenue to 2%. Although payment adjustments began to occur in 2013, we considered the start date to be July 2011, the first period in which hospital performance on quality measures was subject to incentives (with 2013 payment adjustments reflecting 2011 to 2012 performance)</p> <p>Control: no P4P</p>
Outcomes	<p>Clinical process (7 indicators, e.g. patients receiving primary percutaneous coronary intervention received within 90 minutes), patient-experience (8 indicators, e.g. satisfaction with communication with nurses), mortality (30 days), patients who were admitted to the hospital for acute myocardial infarction, heart failure, and pneumonia</p>
Notes	<p>Funding/conflict of interest: Supported by grants from the National Institute on Aging</p> <p>Information for subgroup analysis: teaching status</p> <p>Other comments: -</p>

Risk of bias

Pay for performance for hospitals (Review)

Ryan 2017 (Continued)

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention cannot be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Process outcomes
Blinding of outcome assessment (detection bias, subjective outcomes)	Unclear risk	Patient experience: blinding not specified
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not sufficiently reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Unclear risk	Analyses outcome same as outcome linked to payment
Baseline outcomes similar All outcomes	Low risk	Propensity score matching
Free of contamination	Low risk	Allocation on hospital level
Baseline characteristics similar	Low risk	Propensity score matching

Schuller 2014

Methods	Study design (CBA, ITS): ITS Allocation linked to study or natural experiment: natural experiment Selection of hospitals (intervention/control): NR Unit of allocation (region, hospital, country): NA Nature of desired change: initiation of nonpayment Data source: 2005-2009 Nationwide Inpatient Sample (NIS) datasets Unit of analyses: patients Number of measurements (before, transition, after, unit [e.g. years]): 15/-/5 (quarter) Statistical analyses:
---------	---

Schuller 2014 (Continued)

- Method: Poisson regression
- Adjustment factors: bed size, teaching status, control and ownership, location, region, age, primary payer

Participants	<p>Country/region/setting: USA/all/NR</p> <p>Health system characteristics: NR</p> <p>Number of hospitals included in the analysis (total): 1050</p> <p>Characteristics of hospitals (before [whole population], intervention/control):</p> <ul style="list-style-type: none">• Number of beds: NR• Number of patients: NR• Number of wards: NR• Departments: NR• Owner: NR• Teaching status: NR• Location: NR <p>Number of patients included in the analysis (total): 40,082,431 (discharges)</p> <p>Characteristics of patients (before [whole population], intervention/control):</p> <ul style="list-style-type: none">• Age: NR• Gender: NR• Casemix: NR• Indications: - <p>Existing/other quality programs: -</p> <p>Other relevant context information: none</p>
Interventions	<p>Nonpayment (no reimbursement) for hospital acquired catheter-associated urinary tract infections</p> <p>Control:</p> <p>No P4P</p>
Outcomes	<p>Catheter-associated urinary tract infections (measurement NR)</p>
Notes	<p>Funding/conflict of interest: NR</p> <p>Information for subgroup analysis: -</p> <p>Other comments: -</p>
<p>Risk of bias</p>	
Bias	<p>Authors' judgement</p> <p>Support for judgement</p>
Intervention independent (ITS)	<p>Unclear risk</p> <p>Not enough information provided</p>
Shape of effect pre-specified (ITS)	<p>Low risk</p> <p>Rational for shape is given</p>
Unlikely to affect data collection (ITS) All outcomes	<p>High risk</p> <p>Analysed outcome same as outcome linked to payment</p>

Pay for performance for hospitals (Review)

Schuller 2014 (Continued)

Knowledge of the allocated interventions, Blinding (ITS, objective outcomes)	Unclear risk	Not specified
Incomplete outcome data addressed (ITS) All outcomes	Low risk	Catheter-associated urinary tract infections
Free of selective reporting (ITS)	Low risk	No indication for selective reporting
Free of other bias (ITS)	Low risk	No evidence for other bias

Shih 2014

Methods	<p>Study design (CBA, ITS): CBA</p> <p>Allocation linked to study or natural experiment: linked to study</p> <p>Selection of hospitals (intervention/control):</p> <ul style="list-style-type: none"> Intervention: hospitals participating the Premier Hospital Quality Incentive Demonstration (HQID) Control: hospitals not participating the Premier Hospital Quality Incentive Demonstration (HQID) <p>Unit of allocation (region, hospital, country): hospitals</p> <p>Nature of desired change: modification of P4P</p> <p>Data source: State Inpatient Database. Data on hospital characteristics was obtained from the American Hospital Association Annual Survey</p> <p>Unit of analyses:</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): 1/-/1 (45/38 months)</p> <p>Statistical analyses:</p> <ul style="list-style-type: none"> Method: difference-in-difference (logistic regression) Adjustment factors: comorbid diseases
Participants	<p>Country/region/setting: USA /12 geographically dispersed states/ hospitals</p> <p>Health system characteristics: NR</p> <p>Number of hospitals included in the analysis (intervention/control): 44/321 (CABG); 93/1046 (hip and knee replacement)</p> <p>Characteristics of hospitals (intervention/control):</p> <p>Coronary artery bypass surgery</p> <ul style="list-style-type: none"> Number of beds (measure NR): 624/563 Number of patients: NR Number of wards: NR Departments: NR Owner: NR Teaching status (teaching): 30.6%/24.6% Location (urban): 97.2%/96.8%

Shih 2014 (Continued)

Hip and knee replacement

- Number of beds (measure NR): 484/399
- Number of patients: NR
- Number of wards: NR
- Departments: NR
- Owner: NR
- Teaching status (teaching): 14.8%/16.5%
- Location (urban): 93.5%/93.4% (hip and knee replacement)

Number of patients included in the analysis (intervention/control): 30,794 / 25,974 (CABG admissions); 23,307 / 17,974 (hip and knee replacement admissions)

Characteristics of patients (intervention/control):

Coronary artery bypass surgery

- Age: 72.7%/72.7%
- Gender: 31.4%/31.6%
- Casemix: NR

Hip and knee replacement

- Age: 73.2%/73.4%
- Gender: 65.5%/65.1%
- Casemix: NR

Existing/other quality programs: NR

Other relevant context information: none

Interventions	<p>Phase 1 (before): top 20% of hospitals received 1 to 2% bonuses in Medicare reimbursements</p> <p>Phase 2 (after): financial bonuses were additionally given to hospitals that significantly improved on their performance. Hospitals could now qualify for bonuses in 3 ways: (1) performing in the top 20% of hospitals ("Top Performance Award") (2) performing above the median level of performance in the current year and ranking in the top 20% in terms of improvement ("Improvement Award") and (3) performing above the median level of performance for a composite quality score benchmark from 2 years prior ("Attainment Award"). Over the 6 years of the demonstration, CMS awarded more than USD 60 million in financial bonuses with almost USD 12 million in incentive payments in the final year</p> <p>Control: no P4P</p>
Outcomes	Inpatient mortality (30 days), inpatient complication (ICD-Codes, follow-up not specified), serious inpatient complications (ICD-Codes, follow-up not specified)
Notes	<p>Funding/conflict of interest: study was supported by grants to Dr. Shih from the National Institutes of Health and Drs. Dimick, Nicholas, and Birkmeyer from the National Institute of Aging. Dr. Dimick and Dr. Nicholas are also supported by career development awards from the Agency for Healthcare Research and Quality and the National Institute on Aging (K01AG041763)</p> <p>Information for subgroup analysis: -</p> <p>Other comments: -</p>
Risk of bias	
Bias	Authors' judgement Support for judgement

Shih 2014 (Continued)

Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention can not be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Mortality, inpatient complication
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Low risk	No evidence of other risk of bias
Baseline outcomes similar All outcomes	Unclear risk	Baseline outcome not reported for control group
Free of contamination	Low risk	It is unlikely that the control received the intervention
Baseline characteristics similar	Low risk	Adjusted for in analysis

Sutton 2012

Methods	<p>Study design (CBA, ITS): CBA</p> <p>Allocation linked to study or natural experiment: natural experiment</p> <p>Selection of hospitals (intervention/control): hospitals treating 100 patients for the condition during the before and the after period</p> <p>- Intervention: all National Health Service (NHS) hospitals in the northwest region of England</p> <p>- Control: all NHS hospitals in other regions of England</p> <p>Unit of allocation (region, hospital, country): region</p> <p>Nature of desired change: initiation of P4P</p> <p>Data source: national Hospital Episode Statistics from the NHS Information Centre for Health and Social Care. Hospital characteristics were obtained from the websites of national regulators and the NHS Information Centre</p> <p>Unit of analyses: hospitals</p> <p>Number of measurements (before, transition, after, unit [e.g. years]): 6/-/6 (quarter)</p> <p>Statistical analyses:</p>
---------	--

Sutton 2012 (Continued)

- Method: difference-in-differences (logistic regression)
- Adjustment factors: sex, age, the primary diagnosis code; 31 coexisting conditions, type of admission (emergency or transfer from another hospital), location from which the patient was admitted

Participants

Country/region/setting: England/see selection of hospitals/general hospitals

Health system characteristics: national health system

Number of hospitals included in the analysis (intervention/control): 24/132

Characteristics of hospitals (intervention/control):

- Number of beds:
 - * - large (number not specified): 7 (29%) / 38 (29%)
 - * - medium (number not specified): 8 (33%) / 44 (33%)
 - * - small (number not specified): 4 (17%) / 27 (20%)
- Number of patients: NR
- Number of wards: NR
- Departments: NR
- Owner: NA
- Teaching status (teaching or specialist): 5 (21%)/23 (17%)
- Location: NA

Number of patients included in the analysis (intervention/control): 38,823 / 206,364 (AMI); 30,918 / 170,085 (heart failure); 64,694/345,690 (pneumonia)

Characteristics of patients (intervention/control):

- AMI
 - * age (mean, before, after): 70.2 (before), 70.2 (after) / 70.3 (before), 70.7 (after)
- Heart failure
 - * age (mean, before, after): 75.9 (before), 76.6 (after) / 77.5 (before), 78.1 (after)
- Pneumonia
 - * age (mean, before, after): 71.8 (before), 72.4 (after) / 72.4 (before), 73.1 (after)

Existing/other quality programs: Quality improvement was supported by other mechanisms, including feedback of data from Premier on performance, centralized support to ensure standardization of data collection, and a range of quality-improvement activities within hospitals. Regular shared-learning events for hospitals involved in the program. Composite results were publicly reported on a dedicated website

Other relevant context information: none

Interventions

Hospitals that reported quality scores in the top quartile received a bonus payment equal to 4% of the revenue that they received under the national tariff for the associated activity. For hospitals in the second quartile, the bonus was 2%. For the next 6 months, the reward system changed so that bonuses could be earned on the basis of 3 criteria. Hospitals were awarded an “attainment” bonus if their achievement in the second year exceeded the median achievement level from the first year, an “improvement” bonus if their increase in achievement from the first year was in the top quartile of increases in achievement from the first year, and an “achievement” bonus if their level of achievement in the second year was in the top or second quartile of achievement levels in the second year. Hospitals could earn all 3 bonuses and had to achieve the “attainment” bonus to be eligible for the “improvement” and “achievement” bonuses. There were no penalties for poor performers at any stage.

Bonuses totaling USD 5 million (GBP 3.2 million) were paid to hospitals at the end of the first year. Bonuses totaling USD 2.5 million (GBP 1.6 million) were paid 6 months later. Thereafter, the program was absorbed into a new pay-for-performance program that applied across the whole of England.

Control:

Sutton 2012 (Continued)

No P4P

Outcomes	Mortality (30 days)	
Notes	Funding/conflict of interest: Supported by the National Health Service National Institute for Health Research. Some of the others received grants from public organizations (e.g. NHS). Information for subgroup analysis: Other comments:	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA, intervention can not be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Mortality
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Low risk	No evidence of other risk of bias
Baseline outcomes similar All outcomes	High risk	Outcome measurements differ
Free of contamination	Low risk	It is unlikely that the control received the intervention
Baseline characteristics similar	Low risk	Adjusted for in analysis

Waters 2015

Methods

Study design (CBA, ITS): ITS

Allocation linked to study or natural experiment: natural experiment

Selection of hospitals (intervention/control): nonfederal US hospitals contributing data to the National Database of Nursing Quality Indicators (NDNQI)

Unit of allocation (region, hospital, country): NA

Waters 2015 (Continued)

Nature of desired change: initiation

Data source: National Database of Nursing Quality Indicators (NDNQI), a program of the American Nurses Association. The NDNQI data were combined with American Hospital Association, Medicare Cost Report, and local market data.

Unit of analyses: hospitals

Number of measurements (before, transition, after, unit [e.g. years]): hospital-acquired pressure ulcers 5/-/4; injurious falls 7/-/7; central line-associated bloodstream infections 4/-/8; catheter-associated urinary tract infections 3/-/9 (quarter)

Statistical analyses:

- Method: difference-in difference, negative- and beta-binomial models, random effects model
- Adjustment factors:

Participants

Country/region/setting: USA/whole country/hospitals

Health system characteristics: Medicare

Number of hospitals included in the analysis (total): 1381

Characteristics of hospitals (before):

- Number of beds:
 - * <100: 18.3% (falls); 17.5% (pressure ulcers); 14.9% (central line-associated bloodstream infections); 14.6% (catheter-associated urinary tract Infections)
 - * 100 to 399: 59.2% (falls); 59.7% (pressure ulcers); 63.8% (central line-associated bloodstream infections); 65.1% (catheter-associated urinary tract Infections)
 - * ≥400: 22.5% (falls); 22.8% (pressure ulcers); 21.3% (central line-associated bloodstream infections); 20.3% (catheter-associated urinary tract Infections)
- Number of patients: NR
- Number of wards: NR
- Departments: NR
- Owner: NR
- Teaching status:
 - * teaching: 15.0% (falls); 15.1% (pressure ulcers); 13.6% (central line-associated bloodstream infections); 12.3% (catheter-associated urinary tract Infections)
 - * residency training: 23.2% (falls); 23.4% (pressure ulcers); 21.2% (central line-associated bloodstream infections); 21.5% (catheter-associated urinary tract Infections)
 - * non-teaching: 61.8% (falls); 61.5% (pressure ulcers); 65.2% (central line-associated bloodstream infections); 66.2% (catheter-associated urinary tract Infections)
- Location:
 - * metropolitan: 85.2% (falls); 86.0% (pressure ulcers); 85.6% (central line-associated bloodstream infections); 56.5% (catheter-associated urinary tract Infections)
 - * micropolitan: 11.5% (falls); 11.0% (pressure ulcers); 12.7% (central line-associated bloodstream infections); 13.8% (catheter-associated urinary tract Infections)
 - * rural: 3.3% (falls); 3.0% (pressure ulcers); 1.7% (central line-associated bloodstream infections); 1.8% (catheter-associated urinary tract Infections)

Number of patients included in the analysis (before/after or intervention/control): NR

Characteristics of patients (before [whole population], intervention/control): data were obtained for adult medical, surgical, step-down, and intensive care units

- Age: NR
- Gender: NR
- Casemix: NR
- Indications:

Waters 2015 (Continued)

Existing/other quality programs: NR

Other relevant context information: none

Interventions	Non-payment for Hospital-Acquired Conditions (injury from falls, hospital-acquired pressure ulcers, catheter-associated urinary tract infections, central line-associated bloodstream infections) Control: No P4P
Outcomes	Injury from falls, hospital-acquired pressure ulcers, catheter-associated urinary tract infections, central line-associated bloodstream infections
Notes	Funding/conflict of interest: Dr Waters, Daniels, Bazzoli, Perencevich, Dunton, Staggs, Fareed, and Shorr and Ms Potter were supported by the Agency for Healthcare Research and Quality during the conduct of this study. Drs Waters, Daniels, Dunton, Staggs, and Shorr and Ms Potter were also supported by the National Institute on Aging; no conflict of interest reported. Information for subgroup analysis: no Other comments:-

Risk of bias

Bias	Authors' judgement	Support for judgement
Intervention independent (ITS)	Unclear risk	Not enough information provided
Shape of effect pre-specified (ITS)	Low risk	Rational for shape is given
Unlikely to affect data collection (ITS) All outcomes	High risk	Analysed outcome same as outcome linked to payment
Knowledge of the allocated interventions, Blinding (ITS, objective outcomes)	Low risk	Injury from falls, catheter-associated urinary tract infections, central line-associated bloodstream infections
Knowledge of the allocated interventions, Blinding (ITS, subjective outcomes)	Unclear risk	Hospital-acquired pressure ulcers, blinding not specified
Incomplete outcome data addressed (ITS) All outcomes	Unclear risk	Not enough information
Free of selective reporting (ITS)	Low risk	No indication for selective reporting
Free of other bias (ITS)	Low risk	No evidence for other bias

Werner 2011

Methods	Study design (CBA, ITS): CBA
---------	-------------------------------------

Pay for performance for hospitals (Review)

Werner 2011 (Continued)

Allocation linked to study or natural experiment: linked to study

Selection of hospitals (intervention/control):

- Intervention: hospitals that agreed to participate
- Control: matched group of hospitals not in the project during the demonstration's first 5 years

Unit of allocation (region, hospital, country): hospitals

Nature of desired change: modification of P4P

Data source: Hospital Compare data available on the CMS. Data were supplemented with hospital characteristics from the Medicare Provider of Service File and Impact File. Data on hospital financial status were from Medicare Cost Reports.

Unit of analyses: hospitals

Number of measurements (before, transition, after, unit [e.g. years]): 9/-/11 (quarter)

Statistical analyses:

- Method: propensity-score matching, logistic regression.
- Adjustment factors:
 - * matching criteria: hospital costs, risk-standardized mortality rates, geographic region.
 - * adjustment factors: number of beds, ownership, teaching status, accreditation, nurse-to-bed ratio, percentage of Medicare admissions, urban or rural location, the percentage of a hospital's patient days that are attributable to low-income patients, level of market competition.

Participants

Country/region/setting: USA/NR/acute care hospitals

Health system characteristics: Medicare and Medicaid patients

Number of hospitals included in the analysis (intervention/control): 260/780

Characteristics of hospitals (intervention/control):

- Number of beds (mean): 259/233
- Number of patients: NR
- Number of wards: NR
- Departments: NR
- Owner:
 - * for profit: 1.5%/2.6%
 - * not for profit: 86.9%/ 84.2%
 - * government: 11.6%/13.3%
- Teaching status (medical school affiliation): 57.1%/62.2%
- Location (rural): 18.5/24.8

Number of patients included in the analysis (before/after or intervention/control): NR

Characteristics of patients (intervention/control):

- Age: NR
- Gender: NR
- Casemix (mean): 1.40/1.43
- Indications: acute myocardial infarction, heart failure, and pneumonia

Existing/other quality programs: public reporting

Other relevant context information: none

Interventions

Before: over the first 2 years of the demonstration project, financial bonuses were distributed to the top 20 percent of hospitals.

Werner 2011 (Continued)

After: In the third year, the bonuses continued, and hospitals performing below a threshold level had to pay penalties for their low performance. 2 additional payment incentives were introduced in the fourth year. Hospitals that attained a target performance level (defined as median performance 2 years previously) received an incentive. In addition, of the hospitals attaining that level, those that were in the top 20 percent in terms of improvement received another incentive.

During the demonstration project, the amount of the incentive that hospitals were eligible for was directly proportional to the base Medicare payment that they received for patients treated for each targeted clinical condition. In other words, the more Medicare patients a hospital treated for a targeted condition, the larger the possible incentive for that condition. During the project's first 5 years, CMS paid participating hospitals more than USD 48 million in rewards.

Control:

No P4P

Outcomes	Quality composite score (weighted average across performance measures, where each measure is weighted by the number of people who are eligible for it) <ul style="list-style-type: none">• Acute Myocardial Infarction:<ul style="list-style-type: none">* aspirin at arrival* aspirin prescribed at discharge* ACEI or ARB for LVSD* adult smoking cessation advice/counseling* beta-blocker prescribed at discharge* beta-blocker at arrival* fibrinolytic received within 30 minutes of hospital arrival* primary PCI received within 90 minutes of hospital arrival* risk adjusted inpatient mortality• Heart Failure:<ul style="list-style-type: none">* evaluation of LVS function (LVF Assessment)* ACEI/ARB for LVSD* detailed discharge instructions* adult smoking cessation advice/counseling• Pneumonia:<ul style="list-style-type: none">* oxygenation assessment* initial antibiotic selection for CAP in immunocompetent patients* blood cultures performed in the ED prior to initial antibiotic received in hospital* influenza vaccination* pneumococcal vaccination* adult smoking cessation advice/counseling	
Notes	Funding/conflict of interest: research was supported by a grant from the Agency for Healthcare Research and Quality. Rachel Werner is supported in part by a Department of Veterans Affairs Health Services Research and Development Career Development Award. Information for subgroup analysis: - Other comments: -	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA

Werner 2011 (Continued)

Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	CBA
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Quality composite score
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Low risk	No evidence of other risk of bias
Baseline outcomes similar All outcomes	Low risk	Matched sample
Free of contamination	Low risk	Not reported
Baseline characteristics similar	Low risk	Matched sample

Zuckermann 2016
Methods
Study design (CBA, ITS): CBA

Allocation linked to study or natural experiment: natural experiment

Selection of hospitals (intervention/control):

Intervention: index stays for the 3 conditions targeted by the Hospital Readmissions

Reduction Program (acute myocardial infarction, heart failure, and pneumonia) were identified, using the program's inclusion and exclusion criteria

Unit of allocation (region, hospital, country): individual

Nature of desired change: implementation

Data source: Medicare Part A and Part B claims for fee-for-service beneficiaries 65 years of age or older who were enrolled for 1 year before they had an index hospitalization in an acute care hospital during the period from October 2007 through May 2015

Unit of analyses: individual level

Number of measurements (before, transition, after, unit [e.g. years]): 10/10/11 (quarter)

Statistical analyses:

- Method: linear generalized estimating equation models, controlled interrupted time series models

Zuckermann 2016 (Continued)

- Adjustment factors: age, 31 coexisting medical conditions, and principal discharge diagnosis, time trend

Participants	<p>Country/region/setting: USA/New York State/hospitals</p> <p>Health system characteristics: Medicare</p> <p>Number of hospitals included in the analysis (totally): 3387</p> <p>Characteristics of hospitals (intervention/control):</p> <ul style="list-style-type: none"> • Number of beds: <ul style="list-style-type: none"> * 0 to 250: before 47.4%, after 45.8%/before 42.7%, after 40.5% * 251 to 500: before 32.6%, after 33.7%/before 34.7%, after 35.3% * ≥ 501: before 17.1%, after 18.4%/before 20.0%, after 22.3% • Number of patients: NR • Number of wards: NR • Departments: NR • Owner (for profit): before 16.6%, after 16.6%/ before 16.2%, after 16.3% • Teaching status (teaching): before 47.0%, after 47.7%/ before 51.1%, after 52.7% • Location (urban): before 80.7%, after 83.2%/ before 84.3%, after 87.0% <p>Number of patients included in the analysis (total): 7,175,558 (targeted condition stays)/ 45,495,870 (nontargeted condition stays)</p> <p>Characteristics of patients (intervention/control):</p> <ul style="list-style-type: none"> • Age: <ul style="list-style-type: none"> * 65 to 74: before 29.2%, after 31.9%/before 36.6%, after 39.3% * 75 to 84: before 39.8%, after 35.6%/before 39.8%, after 35.3% * ≥85: before 31.0%, after 32.5%/before 23.6%, after 25.4% • Gender (male): before 45.6%, after 47.7%/before 41.9%, after 43.7% • Casemix: NR • Indications: acute myocardial infarction, heart failure, and pneumonia/other conditions <p>Existing/other quality programs: NR</p> <p>Other relevant context information: none</p>
Interventions	<p>Hospital Readmission Reduction Program penalized hospitals with higher than- expected 30-day readmission rates. In FY 2013 and 2014, the conditions were acute myocardial infarction, heart failure, and pneumonia. Total hip or knee replacement and chronic obstructive pulmonary disease (COPD) were added in FY 2015. Initially, in FY 2013, the maximum penalty was 1% of a hospital's Medicare base diagnosis-related-group (DRG) payments, but the penalty has been increased to 3% for FY 2015 and the years beyond.</p> <p>Control: no penalties</p>
Outcomes	Readmissions within 30 days, observation service use within 30 days
Notes	<p>Funding/conflict of interest: Funding not reported/some of the authors were federal employees.</p> <p>Information for subgroup analysis: -</p> <p>Other comments: -</p>
Risk of bias	
Bias	Authors' judgement Support for judgement

Zuckermann 2016 *(Continued)*

Random sequence generation (selection bias)	High risk	CBA
Allocation concealment (selection bias)	High risk	CBA
Blinding of participants and personnel (performance bias) All outcomes	High risk	Intervention cannot be blinded
Blinding of outcome assessment (detection bias) Objective outcomes	Low risk	Re-admission and use of observational services objective outcome
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not reported
Selective reporting (reporting bias)	Low risk	No indication for selective reporting
Other bias	Low risk	No evidence for other bias
Baseline outcomes similar All outcomes	High risk	Baseline outcome measurements different
Free of contamination	High risk	Allocation on patient level
Baseline characteristics similar	Low risk	

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Atkinson 2010	No ITS and not sufficient data for reanalysis
Averill 2011	Narrative review
Bastian 2016	Cohort study
Berthiaume 2006	Time series, only 2 measurements before and after
Bhattacharyya 2008	Influence of the P4P program not analyzed, but only factors associated with size of bonus
Calikoglu 2012	Cohort study, i.e. no before measurements
Chen 2016	Primary care
Chen 2017	Before-after study. No sufficient data for reanalysis
Collier 2007	Primary care

Pay for performance for hospitals (Review)

Study	Reason for exclusion
Epstein 2014a	Cohort study, no before measure incorporated in the analysis
Epstein 2014b	No primary or secondary outcome reported
Glickman 2007	Cohort study, no before measure incorporated in the analysis
Hart-Hester 2008	No empirical study
Jha 2010	Cohort study, no before measure incorporated in the analysis
Katz 2010	No empirical study
Kim 2015	No finical incentive to increase quality
Kristensen 2016	No control
Lindenauer 2007	Cohort study, no before measure
Padula 2016	Not analyzed as ITS and no reanalysis possible with the available data
Park 2011	No empirical study
Rieger 2009	No empirical study
Ryan 2012b	No relevant outcome reported
Ryan 2014	No intervention phase
Sautter 2007	Qualitative study
Thirukumaran 2017	Cohort study, Difference-in-difference analysis only for hospital load
Vaz 2015	CBA, but no control without/other P4P program

Characteristics of ongoing studies *[ordered by study ID]*

Bawo 2015

Trial name or title	Quality-based pay for performance scheme in Liberia
Methods	CBA
Participants	Hospitals
Interventions	P4P
Outcomes	Mortality, morbidity, readmissions, length of stay, medical errors,
Starting date	Unclear
Contact information	implementationscience.biomedcentral.com/articles/10.1186/s13012-014-0194-9 kleonard@arec.umd.edu

Pay for performance for hospitals (Review)

Bawo 2015 (Continued)

Notes

ADDITIONAL TABLES

Table 1. Description of P4P programs

Design feature	Premier Hospital Quality Incentive Demonstration	Value-Based Purchasing	Non-payment for Hospital-Acquired Conditions	Hospital Readmissions Reduction Program	Advancing Quality Program (period 2)	Financial Incentive to Quality Improvement
Country	USA	USA	USA	USA	England	France
Period (program start)	2003	2013	2007/2008	2012	2008	2012
Type of incentive	Rewards (some hospitals) or penalties (other hospitals)	Rewards or penalties (losing and winning hospitals)	Penalties	Penalties	1. Period: Rewards 2. Period: Penalties	Rewards
Payment type	Absolute achievement compared with the national average (relative performance)	Absolute achievement compared with the national average and baseline performance (relative performance)	Absolute performance	Absolute achievement compared to the national average (relative performance)	1. Phase: Additional payments if a required quality threshold (absolute achievement) was reached 2. Phase: Payments withheld and only paid if a required quality threshold (absolute achievement) was reached	Absolute achievement compared to the median (relative performance)
Payment target	Quality attainment, Substantial improvement	Quality attainment and quality improvement	Quality attainment	Quality attainment	1. Phase: quality improvement 2. Phase: quality attainment	Quality attainment and quality improvement
Bonus/penalty sizes size	2% of Medicare payment (hospitals in the first decile), 1% (hospitals in the second decile), penalties for very low performing hospitals	2% of DRG revenue	Depending on condition	Maximum penalty 3%	1. Phase: GDP 4.8 million 2. Phase: GDP 3.2 million losses in total each year (all hospitals)	Hospitals receive from 0.3% to 0.5% of their annual budget with maximum payments 600,000 EUR per hospital
Outcome (type)	Quality process score	Clinical processes, patient experi-	Hospital acquired condition	Readmissions within 30 days	Clinical process and outcome measures (5 clinical areas), patient	Clinical process measures

Table 1. Description of P4P programs *(Continued)*

linked to P4P		ence, patient outcomes, re- source use	(11 health condi- tions)	of dis- charge (7 health con- ditions)	reported outcomes, patient experience	
Frequency of quality monitor- ing	Annually	Annually	Ongoing	Ongoing	Annually	At end of the study/pilot phase
Frequen- cy of pay- ment	Annually	Annually	Ongoing	Ongoing	Annually	At end of the study/pilot phase
Obligation	Voluntary	Obligatory	Obliga- tory	Obligatory	Voluntary	Voluntary
Coverage	Medicare/Medicaid acute in-patient care for 6 clini- cal conditions	Medicare acute inpatient care	Medicare inpatient care	Medicare inpatient care	Hospitals providing emergency care	Acute care hospi- tals
Budget	Extra budget (USD 12 million in incentive pay- ments in the final year)	Budget neutral	Savings	Savings	Budget neutral	Extra budget

Table 2. Results Premier Hospital Quality Incentive Demonstration (CBA)

Study/Outcome	Intervention			Control			Absolute difference in change or risk ratio (95% CI or P value)	Interactions
	Base-line	Transition	Change	Base-line	Transition	Change		
Grossbart 2006 Quality score (overall, mean)	80.4	- 89.7	9.3	78.9	- 85.6	6.7	2.6 (P < 0.001)	-
Quality score (acute myocardial infarction, mean)	91.1	- 94.2	3.1	87.8	- 90.6	2.9	0.2 (P < 0.730)	-
Quality score (heart failure, mean)	67.8	- 87.0	19.2	73.4	- 84.3	10.9	8.2 (P < 0.001)	-
Jha 2012 Mortality (all conditions, %)	12.33	- 11.82	-0.04*	12.40	- 11.74	-0.04*	0.01 (95% CI -0.02 to 0.01)	Level of financial incentive
Mortality (acute myocardial infarction, %)	17.32	- 15.67	-0.11*	17.42	- 15.85	-0.09*	0.02 (95% CI -0.05 to 0.01)	-
Mortality (congestive heart failure, %)	10.68	- 11.13	-0.01*	10.61	- 10.92	-0.01*	0.00 (95% CI -0.02 to 0.02)	-
Mortality (pneumonia, %)	12.87	- 11.71	-0.07*	13.13	- 11.85	-0.06*	0.01 (95% CI -0.03 to 0.02)	-
Mortality (coronary-artery bypass grafting, %)	3.91	- 4.12	-0.03*	3.62	- 3.34	-0.02*	0.01 (95% CI -0.03 to 0.02)	-
Kruse 2012 Hospital costs (mean per admission)	16,982	- -	-	15,932	- -	-	NR (P > 0.05)	-
Ryan 2009 Mortality (acute myocardial infarction, %)	14.4	- -	-	16.9	- -	-	-1.3 (P > 0.1)	-
Mortality (congestive heart failure, %)	9.0	- -	-	9.7	- -	-	1.9 (P > 0.1)	-
Mortality (pneumonia, %)	11.0	- -	-	11.1	- -	-	0.9 (P > 0.1)	-
Mortality (coronary artery bypass grafting, %)	3.7	- -	-	2.8	- -	-	5.0 (P > 0.1)	-

Table 2. Results Premier Hospital Quality Incentive Demonstration (CBA) (Continued)

Hospital costs (acute myocardial infarction, thousands of U.S. dollars, mean)	25.1	-	-	-	24.6	-	-	-	-0.9 (P > 0.1)	-
Hospital costs (congestive heart failure, thousands of U.S. dollars, mean)	13.4	-	-	-	12.4	-	-	-	0.9 (P > 0.1)	-
Hospital costs (pneumonia, thousands of U.S. dollars, mean)	9.8	-	-	-	9.7	-	-	-	-0.3 (P > 0.1)	-
Hospital costs (coronary-artery bypass grafting, thousands of U.S. dollars, mean)	37.3	-	-	-	34.7	-	-	-	-0.8 (P > 0.1)	-
Ryan Received CBAG (white vs non-white, %)	2.3	-	2.0	0.3	2.4	-	2.6	-0.2	-0.6 (P > 0.1)	-
2010 Received CBAG (White vs other ethnic group, %)	-1.6	-	-0.8	-0.8	-1.5	-	0.7	-2.2	-1.5 (P < 0.1)	-
Received CBAG (White vs Black, %)	2.6	-	3.1	-0.5	3.7	-	3.4	0.3	-0.1 (P > 0.1)	-
Received CBAG (White vs Hispanic, %)	0.2	-	0.3	0.1	0.3	-	0.4	-0.1	-0.3 (P > 0.1)	-
Ryan Composite quality score (pneumonia, mean)	89.2	-	NR	NR	88.4	-	NR	NR	-0.67 (P > 0.1)	-
2011 Composite quality score (surgical site infection, mean)	86.3	-	NR	NR	81.1	-	NR	NR	-0.12 (P > 0.1)	-
Ryan Composite quality score (heart attack, mean annual improvement)	1.17	-	1.57	0.4	0.37	-	1.27	0.9	-0.50 (95% CI -1.12 to 0.13)	Effect does not vary across type of ownership, hospital size, region, teaching
2012 Composite quality score (heart failure, mean annual improvement)	2.36	-	4.38	2.02	0.33	-	3.80	3.47	-1.45 (95% CI -2.60 to -0.30)	
Composite quality score (pneumonia, mean annual improvement)	3.11	-	3.57	0.46	1.61	-	2.97	1.36	-0.91 (95% CI -1.57 to -0.25)	

Table 2. Results Premier Hospital Quality Incentive Demonstration (CBA) (Continued)

									sta- tus
Shih Mortality (CABG,%)	3.1	- 2.4	0.70 (95% CI 0.66 to 0.75)†	-	- -	-	OR 1.09 (95% CI 0.90 to 1.32)	-	
2014									
Adverse clinical events (CABG, %)	21.7	- 23.2	1.01 (95% CI 0.94 to 1.08)†	-	- -	-	OR 1.13 (95% CI 0.98 to 1.29)	-	
Serious adverse clinical events (CABG, %)	13.5	- 13.6	0.96 (95% CI 0.92 to 1.01)†	-	- -	-	OR 1.06 (95% CI 0.95 to 1.19)	-	
Mortality (joint replacement,%)	0.2	- 0.2	0.78 (95% CI 0.61 to 1.00)†	-	- -	-	OR 0.85 (95% CI 0.54 to 1.32)	-	
Adverse clinical events (joint replacement, %)	4.2	- 4.2	0.89 (95% CI 0.84 to 0.95)†	-	- -	-	OR 1.11 (95% CI 0.91 to 1.36)	-	
Serious adverse clinical events (joint replacement, %)	2.8	- 2.6	0.79 (95% CI 0.74 to 0.84)†	-	- -	-	OR 1.12 (95% CI 0.95 to 1.31)	-	
Werner Quality score (all conditions, phase 1 vs. phase 2)	84.70	- 90.86	6.164	82.18	- 88.19	6.01	0.153	-	
2011 Quality score (all conditions, phase 2 vs. phase 3)	90.86	- 93.88	3.02	88.19	- 91.99	3.80	-0.778	-	
At last observation there was no statistically significant difference between the 2 groups (author statement)									

*change per quarter; †relative change
NR: not reported; OR: odds ratio

Table 3. Results Value-Based Purchasing (CBA)

Study	Outcome	Intervention			Control			Absolute difference in change or risk ratio (95% CI or P value)	Interac- tions
		Base- line	TraAf- si- ter tion	Change	Base- line	TraAf- si- ter tion	Change		
Ryan 2015	Process performance (%)	89.5	-	NR	89.0	-	NR	-0.51 (95% CI -1.37 to 0.34)	-
	Patient experience (%)	68.6	-	NR	68.7	-	NR	-0.30 (95% CI -0.79 to 0.19)	-
Ryan 2017	Clinical-process composite (standardized)	-0.32	-	NR	-0.33	-	NR	0.079 (95% CI -0.140 to 0.299)	No evi- dence that teaching status, hospital size modi- fied the ef- fect
	Patient-experience composite (standardized)	-0.01	-	NR	0.07	-	NR	-0.09 (95% CI -0.31 to 0.12)	
	Mortality (myocardial infarction, %)	16.07	-	NR	16.04	-	NR	-1.474 -0.282 (95% CI -1.72 to 1.15)	
	Mortality (heart failure, %)	11.49	-	NR	11.53	-	NR	-0.212 (95% CI -0.53 to 0.11)	
	Mortality (pneumonia, %)	11.86	-	NR	11.87	-	NR	-0.431 (95% CI -0.71 to -0.15)	
Figure 10a 2016	Mortality (individual level analysis, quarterly changes %)	-0.13	-	-0.030.10	-0.09	-	-0.020.07	0.01 (P = 0.12)	-
	Mortality (hospital level analysis, quarterly changes %)	-0.13	-	-0.030.10	-0.14	-	-0.010.13	-0.03 (95% CI -0.08 to 0.13)	-

NR: not reported

Table 4. Results Non-payment for Hospital-Acquired Conditions (CBA)

Study	Outcome	Intervention			Control			Absolute difference in change or risk ratio (95% CI or P value)	In- ter- ac- tions
		Base- line	Tran- si- tion	Af- ter	Change	Base- line	Tran- si- tion	Af- ter	
									-

Table 4. Results Non-payment for Hospital-Acquired Conditions (CBA) *(Continued)*

Kwong 2017	Surgical site infections (no per 1000)	7.0	-	5.2	-1.8	5.9	-	4.9	-1.0	-0.8 (not estimated)	-
	Surgical site infections (RR)	NR	-	NR	0.7	NR	-	NR	0.8	0.9 (95% CI 0.8 to 1.1)	-

NR: not reported; RR: risk ratio

Table 5. Results Non-payment for Hospital-Acquired Conditions (ITS)

Study	Outcome	Before (for each measure- ment point or trend)	Tran- sition (for each measure- ment point)	After (for each measure- ment point or trend)	Change from base- line, 95% CI or P value [#]	Interac- tions
Kawai 2015	Vascular catheter-associated infections (relative change in trend per quarter, odds ratio)	1.17	0.75	0.98	0.84 (95% CI 0.79 to 0.88)	Size, own- ership
	Catheter-associated urinary tract infec- tions (relative change in trend per quar- ter, odds ratio)	1.19	0.87	0.99	0.83 (95% CI 0.81 to 0.85)	
Lee 2012	Central catheter-associated blood- stream infections (slope of incidence rate ratio)	0.95	NR	0.95	1.00* (95% CI 0.97 to 1.03)	Hospi- tal size, teaching status,
	Catheter-associated urinary tract infec- tions (slope of incidence rate ratio)	0.96	NR	0.99	1.03* (95% CI 1.00 to 1.07)	and type of own- ership, monitor- ing, were not associat- ed with a differ- ential re- sponse (test for inter- action: p≥0.05)
Morgan 2012	Antimicrobial use (relative change %)	0.30	NR	-1.24	P < 0.001	-
Padula[#] 2015	Hospital acquired pressure ulcers (inci- dence rate per 1,000 patients, mean)	10.133	NR	1.204	-8.929	-
	Hospital acquired pressure ulcers (slope for incidence rate per 1,000 patients)	-1.285	NR	-0.084	1.201	-
	Hospital acquired pressure ulcers (lev- el effect for incidence rate per 1,000 pa- tients, first quarter post intervention)	NR	NR	NR	-5.77 (95% CI -2.65 to 8.89)	-
	Hospital acquired pressure ulcers (lev- el effect for incidence rate per 1,000 pa- tients, 5th quarter post intervention)	NR	NR	NR	-1.02 (95% CI -4.13 to 2.09)	-
	Hospital acquired pressure ulcers (lev- el effect for incidence rate [pressure	NR	NR	NR	3.82 (95% CI -2.04 to 9.68)	-

Table 5. Results Non-payment for Hospital-Acquired Conditions (ITS) (Continued)

	ulcers] per 1,000 patients, 9th quarter post intervention)					
	Hospital acquired pressure ulcers (level effect for incidence rate [pressure ulcers] per 1,000 patients, 15th quarter post intervention)	NR	NR	NR	11.331 (95% CI 0.29 to 22.38)	-
Schuller 2014	Catheter-associated urinary tract infections (mean rate, slope)	0.0422	NR	0.0307	-0.012 (statistical uncertainty not reported)	-
	Catheter-associated urinary tract infections (mean rate, intercept)	-15.5609	NR	0.4736 (at P4P implementation)	16.04 (statistical uncertainty not reported)	-
Waters 2015	Pressure ulcers (change in proportion, slope)	0.97 (95% CI 0.96 to 0.99)	NR	0.98 (95% CI 0.96 to 1.00)	1.00 (95% CI 0.98 to 1.03)	-
	Falls (change in proportion, slope)	0.99 (95% CI 0.98 to 0.99)	NR	0.98 (95% CI 0.98 to 0.99)	1.00 (95% CI 0.99 to 1.00)	-
	Central line-associated bloodstream infections (change in proportion, slope)	1.07 (95% CI 1.00 to 1.15)	NR	0.94 (95% CI 0.93 to 0.95)	0.88 (95% CI 0.82 to 0.94)	-
	Catheter-associated urinary tract infections (change in proportion, slope)	1.04 (95% CI 0.99 to 1.10)	NR	0.94 (95% CI 0.93 to 0.95)	0.90 (95% CI 0.85 to 0.95)	-

#study data reanalyzed; #all narrative descriptions according authors; *relative change

NR: not reported

Table 6. Results Hospital Readmissions Reduction Program (CBA)

Study	Outcome	Intervention			Control			Absolute difference in change or risk ratio (95% CI or P value)	Interactions
		Baseline	Transition	After	Change	Baseline	Transition	After	
Melior 2017	Readmission (30 days, myocardial infarction, percent point)	NR	NR	NR	NR NR	NR	NR	NR 0.011 (P > 0.1)	-
	Readmission (30 days, heart failure, percent point)	NR	NR	NR	NR NR	NR	NR	NR 0.002 (P > 0.1)	-
	Readmission (30 days, pneumonia, percent point)	NR	NR	NR	NR NR	NR	NR	NR 0.007 (P > 0.1)	-
De-sai 2016	Readmission (30 days, myocardial infarction, annually change rate %)	0.15 (95% CI -0.11 to 0.40)	-0.49 (95% CI -0.81 to -0.16)	0.09 (95% CI -0.18 to 0.35)	NR -0.59 (95% CI -0.95 to -0.22)	0.48 (95% CI 0.01 to 0.95)	0.06 (95% CI -0.33 to 0.45)	NR NR	-
	Readmission (30 days, heart failure, annually change rate %)	0.10 (95% CI -0.12 to 0.32)	-0.90 (95% CI -1.18 to -0.62)	0.72 (95% CI 0.49 to 0.95)	NR -0.26 (95% CI -0.56 to 0.04)	0.08 (95% CI -0.30 to 0.46)	0.14 (95% CI -0.17 to 0.46)	NR NR	-
	Readmission (30 days, pneumonia, annually change rate %)	0.37 (95% CI 0.10 to 0.64)	-0.57 (95% CI -0.92 to -0.23)	0.05 (95% CI -0.24 to 0.33)	NR -0.12 (95% CI -0.44 to 0.19)	0.53 (95% CI 0.13 to 0.93)	-0.52 (95% CI -0.86 to -0.19)	NR NR	-
Mc-Garry 2016	Readmission (30 days, odds ratio year 2)	NR	NR	NR	NR NR	NR	NR	NR 1.01 (95% CI 0.99 to 1.03)	-
	ED visit (30 days, odds ratio year 2)	NR	NR	NR	NR NR	NR	NR	NR 1.07 (95% CI 1.04 to 1.11)	-
Zucker-mann 2016	Readmission (30 days, %, slope)	-0.017	-0.103	-0.005	0.097 0.008	-0.061	-0.004	0.057 0.032* (95% CI -0.041 to -0.024)	-
	Observational services (30 days, %, slope)	0.020	0.025	0.033	0.008 0.021	0.021	0.023	0.002 0.005* (95% CI < 0.000 to 0.009)	-

NR: not reported

*Change from pre to transition

Table 7. Results Hospital Readmission Reduction Program (ITS)

Study	Out- come	Before (for each measurement point or trend)	Transition (for each measure- ment point)	After (for each mea- surement point or trend)	Change from base- line, 95% CI or P val- ue [#]	Interac- tions
Ibrahim 2017	Read- mission (30 days, % slope)	-0.068	-0.089	-0.098	0.03 (P < 0.001)	-
	Length of stay (% slope)	-0.028	-0.023	-0.037	NR	

[#]all narrative descriptions according authors; ^{*}relative change

Table 8. Results Advancing Quality Program (CBA)

Study Outcome		Intervention			Control			Absolute difference in change or risk ratio (95% CI or P value)			Interactions
		Base-line	Tran-si-tion	After	Change	Base-line	Tran-si-tion	After	Change		
Kris-tensen 2014	Mortality (baseline vs. period 1, %)	20.5	-	18.8	-1.7	18.9	-	18.1	-0.8	-0.9 (95% CI -1.3 to -0.4)	-
	Mortality (baseline vs. period 2, %)	20.5	-	17.2	-3.3	18.9	-	15.7	-3.2	-0.1 (95% CI -0.6 to -0.3)	-
	Mortality (period 1 vs. period 2, %)	18.8	-	17.2	-1.6	18.1	-	15.7	-2.4	0.7 (95% CI 0.3 to 1.2)	-
Sut-ton 2012	Mortality (overall, %)	21.9	-	20.1	-1.8	20.2	-	19.3	-0.9	-0.9 (95% CI -1.7 to -0.1)	-
	Mortality (AMI, %)	12.1	-	10.7	-1.4	11.3	-	10.4	-1.0	-0.4 (95% CI -1.3 to 0.6)	-
	Mortality (heart failure, %)	18.8	-	17.5	-1.3	16.9	-	15.8	-1.1	-0.4 (95% CI -1.5 to 0.7)	-
	Mortality (pneumonia, %)	29.4	-	27.0	-2.4	27.1	-	26.3	-0.7	-1.5 (95% CI -2.5 to -0.5)	-

Table 9. Results Financial Incentive to Quality Improvement (CBA)

Study	Outcome	Intervention				Control				Absolute difference in change or risk ratio (95% CI or P value)	Interactions
		Base-line	Transition	After	Change	Base-line	Transition	After	Change		
Lalloué 2017	Process quality score (mean)	44.3	-	57.4	13.1	48.3	-	58.6	10.3	4.07 (95% CI -1.04 to 9.17)	-

APPENDICES

Appendix 1. Search strategies

Medline (OVID)

1946 to present

No.	Search terms	Results
1	(pay* adj3 performance).ti.	1068
2	reimbursement, incentive/	3962
3	value-based purchasing/	683
4	(pay* adj3 performance).ti,ab,kf.	2340
5	(nonpayment? or non-payment?).ti,ab,kf.	146
6	p4p.ti,ab,kf.	461
7	(pay* adj3 quality).ti,ab,kf.	854
8	((result? based or performance or output based or out put based or quality based or value based) adj3 (pay* or fee? or incentiv* or remunerat* or reimburs* or compensat* or purchas*)).ti,ab,kf.	4917
9	((payment or financial or monetary) adj (reward* or bonus* or incentiv* or malus* or penalt*)).ti,ab,kf.	6659
10	bonus payment?.ti,ab,kf.	82
11	((target or targets or targeted) adj3 (pay* or reward*)).ti,ab,kf.	482
12	or/2-11	15172
13	hospital*.ti,ab,hw.	1408885
14	units.hw.	98296
15	13 or 14	1461268
16	12 and 15	3519
17	randomized controlled trial.pt.	462868
18	controlled clinical trial.pt.	92461
19	multicenter study.pt.	235037
20	pragmatic clinical trial.pt.	792
21	(randomis* or randomiz* or randomly).ti,ab.	777029
22	groups.ab.	1807519

(Continued)

23	(trial or multicenter or multi center or multicentre or multi centre).ti.	217320
24	(intervention? or effect? or impact? or controlled or control group? or (before adj5 after) or (pre adj5 post) or ((pretest or pre test) and (posttest or post test)) or quasiexperiment* or quasi experiment* or pseudo experiment* or pseudoexperiment* or evaluat* or time series or time point? or repeated measur*).ti,ab.	8495892
25	non-randomized controlled trials as topic/	360
26	interrupted time series analysis/	440
27	controlled before-after studies/	331
28	or/17-27	9482567
29	exp animals/	21595558
30	humans/	17128148
31	29 not (29 and 30)	4467410
32	review.pt.	2394635
33	meta analysis.pt.	89489
34	news.pt.	190389
35	comment.pt.	721830
36	editorial.pt.	461432
37	cochrane database of systematic reviews.jn.	13663
38	comment on.cm.	721826
39	(systematic review or literature review).ti.	113546
40	or/31-39	7930775
41	28 not 40	6640305
42	(1 or 16) and 41	1654

Embase (OVID)

1974 to present

No.	Search terms	Results
1	(pay* adj3 performance).ti.	1226
2	(pay* adj3 performance).ti,ab,kw.	2955

Pay for performance for hospitals (Review)

Copyright © 2019 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

(Continued)

3	(nonpayment? or non-payment?).ti,ab,kw.	183
4	p4p.ti,ab,kw.	542
5	(pay* adj3 quality).ti,ab,kw.	1050
6	((result? based or performance or output based or out put based or quality based or value based) adj3 (pay* or fee? or incentiv* or remunerat* or reimburs* or compensat* or purchas*).ti,ab,kf.	5718
7	((payment or financial or monetary) adj (reward* or bonus* or incentiv* or malus* or penalt*).ti,ab,kf.	8282
8	bonus payment?.ti,ab,kw.	97
9	((target or targets or targeted) adj3 (pay* or reward*).ti,ab,kw.	615
10	or/2-9	15245
11	hospital*.ti,ab,hw.	2219387
12	unit?.hw.	208919
13	11 or 12	2341144
14	10 and 13	3954
15	1 or 14	4823
16	randomized controlled trial/	507203
17	controlled clinical trial/	460135
18	quasi experimental study/	4693
19	pretest posttest control group design/	343
20	time series analysis/	20928
21	experimental design/	15590
22	multicenter study/	188692
23	(randomis* or randomiz* or randomly).ti,ab.	1076344
24	groups.ab.	2467261
25	(trial or multicentre or multicenter or multi centre or multi center).ti.	303657
26	(intervention? or effect? or impact? or controlled or control group? or (before adj5 after) or (pre adj5 post) or ((pretest or pre test) and (posttest or post test)) or quasiexperiment* or quasi experiment* or pseudo experiment* or pseudoexperiment* or evaluat* or time series or time point? or repeated measur*).ti,ab.	10886881
27	or/16-26	12141965

(Continued)

28	(systematic review or literature review).ti.	134424
29	"cochrane database of systematic reviews".jn.	12402
30	exp animals/ or exp invertebrate/ or animal experiment/ or animal model/ or animal tis- sue/ or animal cell/ or nonhuman/	26193669
31	human/ or normal human/ or human cell/	19820678
32	30 not (30 and 31)	6422104
33	28 or 29 or 32	6567641
34	27 not 33	9267909
35	15 and 34	2545

The Cochrane Library (Wiley)

No.	Search terms	Results
#1	(pay* near/3 performance):ti	42
#2	[mh "reimbursement, incentive"]	94
#3	[mh "value-based purchasing"]	2
#4	(pay* near/3 performance):ti,ab	97
#5	(nonpayment? or non-payment?):ti,ab	0
#6	p4p:ti,ab	29
#7	(pay* near/3 quality):ti,ab	34
#8	((result? based or performance or output based or out put based or quality based or value based) near/3 (pay* or fee? or incentiv* or remunerat* or reimburs* or compensat* or pur- chas*)):ti,ab	234
#9	((payment or financial or monetary) next (reward* or bonus* or incentiv* or malus* or penalt*)):ti,ab	1058
#10	bonus next payment?:ti,ab	8
#11	((target or targets or targeted) near/3 (pay* or reward*)):ti,ab	30
#12	{or #2-#11}	1330
#13	hospital*:ti,ab,kw	122790
#14	units:kw	4051

(Continued)

#15	#13 or #14	124729
#16	#12 and #15	190
#17	#1 or #16	224

WHO International Clinical Trials Registry Platform (ICTRP)

pay for performance AND hospital
 payment AND hospital
 nonpayment AND hospital
 p4p AND hospital
 reimbursement AND hospital

ClinicalTrials.gov

hospital AND (pay for performance OR payment OR nonpayment OR p4p OR reimbursement)

Appendix 2. Evidence profile

Premier Hospital Quality Incentive Demonstration compared to Capitation without P4P for Hospitals

Certainty assessment							Summary of findings		
Nº of participants (studies) Follow-up	Risk of bias	Incon-sistency	Indirect-ness	Im-pre-ci-sion	Pub-lica-tion bias	Overall cer-tainty of evi-dence	Study event rates (%)		Impact
							With Capitation without P4P	With Premier Hospi-tal Quality Incentive Demonstration	
Mortality (follow up: 30 days)									
5126 (3 observational studies)	seri-ous	not seri-ous	not seri-ous	seri-ous	none	⊕○○○ VERY LOW	All studies showed a reduction in mortality in the hospitals with P4P but this was throughout very small		
Aderse clinical events (follow up: Inpatient)									
1139 (1 observational study)	seri-ous	not seri-ous	not seri-ous	seri-ous	none	⊕○○○ VERY LOW	P4P reduced adverse clinical events slightly		
Quality of care (follow up: not applicable)									
4249 (3 observational studies)	seri-ous	not seri-ous	not seri-ous	seri-ous	none	⊕○○○ VERY LOW	P4P improved process quality slightly		
Equity (differences between ethnic groups in receiving recommended care) (follow up: not applicable)									
1063 (1 observational study)	seri-ous	not seri-ous	not seri-ous	seri-ous	none	⊕○○○ VERY LOW	P4P increased differences in recommended care slightly		
Cost (follow up: not applicable)									
1392 (2 observational studies)	seri-ous	not seri-ous	not seri-ous	very seri-ous	none	⊕○○○ VERY LOW	P4P had no impact on hospital costs.		

Values based Purchasing compared to Capitation without P4P for Hospitals

Certainty assessment							Summary of findings		
Nº of participants (studies) Follow-up	Risk of bias	Incon-sistency	Indirect-ness	Im-pre-	Pub-lica-	Overall cer-tainty of evi-dence	Study event rates (%)		Impact

(Continued)

							ci- sion	tion bias	With Capitation without P4P	With Values based Pur- chasing
Mortality (follow up: 30 days)										
6248 (2 observational studies)	seri- ous	not seri- ous	not seri- ous	seri- ous	none	⊕○○○ VERY LOW	Very small reduction of mortality in hospitals with P4P			
Process quality score (follow up: not applicable)										
5358 (2 observational studies)	seri- ous	not seri- ous	not seri- ous	seri- ous	none	⊕○○○ VERY LOW	P4P improved process quality slightly			
Non-payment for hospital-acquired conditions program compared to Capitation without P4P for Hospitals										
Bibliography:										
Certainty assessment							Summary of findings			
Nº of participants (studies) Follow-up	Risk of bias	Incon- sistency	Indirect- ness	Im- pre- ci- sion	Pub- lica- tion bias	Overall cer- tainty of evi- dence	Study event rates (%)		Impact	
							With Capitation without P4P	With Non-payment for hospital-acquired con- ditions program		
Clinical adverse events (follow up: in hospital)										
3568 (6 observational studies)	seri- ous	not seri- ous	not seri- ous	not seri- ous	none	⊕○○○ VERY LOW	for hospital acquired conditions reduced clinical adverse events			
Advancing Quality program compared to Capitation without P4P for Hospitals										
Certainty assessment							Summary of findings			
Nº of participants (studies) Follow-up	Risk of bias	Incon- sistency	Indirect- ness	Im- pre- ci- sion	Pub- lica- tion bias	Overall cer- tainty of evi- dence	Study event rates (%)		Impact	
							With Capitation without P4P	With Advancing Quality program		
Mortality (follow up: 30 days)

(Continued)

156 (2 observational studies)	seri- ous	not seri- ous	not seri- ous	seri- ous	none	⊕○○○ VERY LOW	Short term reduction in mortality that could not be sustained
----------------------------------	--------------	------------------	------------------	--------------	------	------------------	---

Financial Incentive to Quality Improvement compared to Capitation without P4P for Hospitals

Certainty assessment							Summary of findings		
N° of participants (studies) Follow-up	Risk of bias	Incon- sistency	Indirect- ness	Im- pre- ci- sion	Pub- lica- tion bias	Overall cer- tainty of evi- dence	Study event rates (%)		Impact
							With Capitation without P4P	With Financial Incen- tive to Quality Im- provement	
Process quality score (follow up: not applicable)									
377 (1 observational study)	seri- ous	not seri- ous	not seri- ous	seri- ous	none	⊕○○○ VERY LOW	Financial Incentive to Quality Improvement improved process quality slightly		

CONTRIBUTIONS OF AUTHORS

Tim Mathes: idea for the review, development of concept, study selection, data extraction, risk of bias assessment, drafting the review

Dawid Pieper: development of concept, study selection, data extraction, risk of bias assessment, revision of review

Johannes Morche: study selection, additional literature searches, data extraction, risk of bias assessment, revision of review

Stephanie Polus: data extraction, risk of bias assessment, revision of review

Thomas Jaschinski: study selection, data extraction, risk of bias assessment, revision of review

Michaela Eikermann: study selection, revision of review

DECLARATIONS OF INTEREST

TM: none known

DP: none known

JM: none known

SP: none known

TJ: none known

ME: none known

SOURCES OF SUPPORT

Internal sources

- None, Other.

External sources

- None, Other.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Initially we planned to analyze basic payment schemes and P4P within the same review because we wanted to consider the effects of P4P in relation to basic payment schemes (see [Discussion](#)). P4P was an add-on to capitation/DRGs in all studies. Therefore, an analysis of the effect on P4P in different basic payment schemes was not possible. Moreover, we identified many more studies than expected. For these reasons we decided to split the review into P4P and the basic payment schemes because this allows us to analyze and discuss the different payment scheme types more comprehensively and increases the readability of the review.

Descriptive data (study and patient characteristics) were extracted by one reviewer and verified by a second reviewer and not as specified in the protocol by two reviewers independently.

We did not perform a meta-analysis because of 'clinical' heterogeneity. Consequently, we could not perform any quantitative analysis that is based on the meta-analysis, including subgroup analysis, sensitivity analysis, imputation of missing data and analysis of funnel plot asymmetry.

We did not apply the double data entry method but all data entries were verified by a second reviewer because most data could not be entered directly into Review Manager 5.

We revised the initial searches completely in order to focus on P4P. We also noted that the first batch of included studies were all indexed in MEDLINE or Embase so we did not search as many sources for subsequent searches. We have reported only the most recent searches for P4P. We revised the inclusion criteria to specify that we considered non-randomized cluster trial designs. We also corrected the inclusion criteria for randomized and non-randomized (controlled) study designs.

Christoph Mosch left the author team. Johannes Morche and Stephanie Polus are new members of the author team.

INDEX TERMS

Medical Subject Headings (MeSH)

*Economics, Hospital; *Quality of Health Care; *Reimbursement, Incentive; Costs and Cost Analysis; Europe; Fee-for-Service Plans; Hospitals; Quality Improvement [economics]; United States

MeSH check words

Humans