



Do penalty-based pay-for-performance programs improve surgical care more effectively than other payment strategies? A systematic review

Kyung Mi Kim^{a,*}, Wendy Max^b, Justin S. White^c, Susan A. Chapman^d, Ulrike Muench^e

^a Clinical Excellence Research Center, School of Medicine, Stanford University, 365 Lasuen St Stanford, CA, 94305, United States

^b Institute for Health & Aging, University of California, 3333 California Street, Suite 340, San Francisco, CA, 94118, United States

^c Philip R. Lee Institute for Health Policy Studies & Department of Epidemiology & Biostatistics, School of Medicine, University of California, 3333 California Street, Suite 265, San Francisco, CA, 94118, United States

^d Department of Social and Behavioral Sciences, School of Nursing, University of California, 3333 California Street, Room 455Q UCSF Box 0612, San Francisco, CA, 94118, United States

^e Department of Social and Behavioral Sciences & Philip R. Lee Institute for Health Policy Studies, School of Nursing, University of California, San Francisco, 3333 California Street, Room 455H UCSF Box 0612, San Francisco, CA, 94118, United States

ARTICLE INFO

Keywords:

Pay-for-performance
Payment strategy
Quality
Cost
Value
Surgical care

ABSTRACT

Background: The aim of this systematic review is to assess if penalty-based pay-for-performance (P4P) programs are more effective in improving quality and cost outcomes compared to two other payment strategies (i.e., rewards and a combination of rewards and penalties) for surgical care in the United States. Penalty-based programs have gained in popularity because of their potential to motivate behavioral change more effectively than reward-based programs to improve quality of care. However, little is known about whether penalties are more effective than other strategies.

Materials and methods: A systematic literature review was conducted according to the PRISMA guideline to identify studies that evaluated the effects of P4P programs on quality and cost outcomes for surgical care. Five databases were used to search studies published from 2003 to March 1, 2020. Studies were selected based on the PRISMA guidelines. Methodological quality of individual studies was assessed based on ROBINS-I with GRADE approach.

Results: This review included 22 studies. Fifteen cross-sectional, 1 prospective cohort, 4 retrospective cohort, and 2 case-control studies were found. We identified 11 unique P4P programs: 5 used rewards, 3 used penalties, and 3 used a combination of rewards and penalties as a payment strategy. Five out of 10 studies reported positive effects of penalty-based programs, whereas evidence from studies evaluating P4P programs with a reward design or combination of rewards and penalties was little or null.

Conclusions: This review highlights that P4P programs with a penalty design could be more effective than programs using rewards or a combination of rewards and penalties to improve quality of surgical care.

1. Introduction

Over 17 million surgical procedures are performed at acute care hospitals in the United States annually, costing the nation approximately \$400 billion per year and accounting for 30% of all health care spending and 50% of hospital costs [1–3]. Surgical complication rates can be as high as 17% depending on the procedure, with one study reporting increases in the costs of a patient stay of \$6139–\$17,850 [4]. One approach to improving surgical care quality and curbing cost are

pay-for-performance (P4P) programs designed with the aim to incentivize providers and hospitals to deliver financially rewarding, high-quality care through linking payment strategies based on rewards or penalties to performance [5–7].

Numerous large-scale P4P programs have been launched by the Centers for Medicare and Medicaid Services (CMS) in inpatient settings over the past decade [8,9] and the number of P4P programs is rapidly expanding. While the greater focus of P4P has been on medical conditions, P4P programs are rapidly expanding in the surgical setting due to

* Corresponding author.

E-mail addresses: kyungkim@stanford.edu, kyungkim@stanford.edu (K.M. Kim), wendy.max@ucsf.edu (W. Max), justin.white@ucsf.edu (J.S. White), susan.chapman@ucsf.edu (S.A. Chapman), ulrike.muench@ucsf.edu (U. Muench).

<https://doi.org/10.1016/j.amsu.2020.11.060>

Received 11 November 2020; Accepted 22 November 2020

Available online 25 November 2020

2049-0801/© 2020 The Authors. Published by Elsevier Ltd on behalf of IJS Publishing Group Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

the strong potential for cost savings on post-operative care (e.g., complications, readmissions). Recently, programs using penalties have gained in popularity [10], because of their potential to affect behavioral change more effectively than reward-based programs. This is due to evidence showing that individuals, and perhaps organizations, tend to be more sensitive to losses than gains [11,12]. However, concerns exist that penalties could harm patients and hospitals by putting financial strain on those hospitals most in need of resources and care improvement [13]. The aim of this systematic review is to assess the current state of evidence on the impact of P4P penalty design on quality and cost outcomes for surgical care compared to two other payment strategies (i. e., reward and a combination of reward and penalty).

Despite the growth in penalty-based P4P programs, few studies have examined the effect of penalty designs compared to other types of payment designs. While several systematic reviews have examined the relationship between P4P program characteristics and their impact on rewards, such as optimal incentive levels and targets, little is known about how payment designs contribute to outcomes. The literature on P4P programs has grown considerably since the latest published systematic review (which searched through 2016) [14], and no studies have specifically focused on surgical care. Although most P4P programs initially targeted medical conditions (e.g., acute myocardial infarction and pneumonia), programs have expanded to target the surgical setting [15]. Given that surgery accounts for the highest inpatient costs with its high spending growth [16] and avoidable complication rates [17], it is important to understand whether penalties are more effective than other strategies in improving surgical care for patients, hospitals and health care delivery more broadly.

2. Methods

2.1. Protocol and registration

We developed a systematic review protocol in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Our study was registered in the Research Registry (registration number: reviewregistry944).

2.2. Systematic literature search

Five databases were used to conduct a systematic review of the effects of P4P payment design on the quality and cost of surgical care: PubMed, Embase, EconLit, the Database of Abstracts of Reviews of Effectiveness, and the Cochrane Database of Systematic Reviews. We supplemented this review with a Google Scholar search, reference list reviews, and a search for selected P4P researchers. Studies on P4P programs in hospitals gained attention over a decade ago [18]; therefore, the search was limited to studies published between 2003 and 2020, conducted on March 1, 2020. Multiple combinations of nine keywords—“pay for performance,” “financial,” “incentive,” “reward,” “bonus,” “penalty,” “reimburse,” “inpatient,” and “surgery”—were used to conduct literature searches in the selected databases. We used “OR” and “AND” as the search terms to include these keywords. We also used an advanced search combining a MeSH major topic, “reimbursement, incentive” and subheading “surgery” as well as a modified strategy employed in a 2014 RAND review of P4P publications [18] and a work by Milstein and Schreyoegg [7]. Using these strategies yielded more than 2800 articles from the five databases. The literature search terms used in each database and the corresponding results in detail are described in Table 1.

Note: DARE = Database of Abstracts of Reviews of Effectiveness. CDSR = Cochrane Database of Systematic Reviews, CMS = Centers for Medicare and Medicaid Services.

2.3. Study selection

Only studies that met all of the following criteria were included: (a) full text available in English, (b) published between 2003 and 2020, (c) evaluated P4P solely implemented in the hospital sector, (d) analyzed the effects of P4P for surgical care, (e) assessed a P4P that used a financial strategy targeting providers and hospitals (as opposed to targeting patients), and (f) demonstrated either quantitative or qualitative empirical analysis, or both.

The study selection process followed protocols under the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). One researcher (K.K.) conducted the database search, and two researchers (K.K. and U.M.) independently reviewed the titles and abstracts of all studies identified in the initial search and selected articles for the primary review following the selection criteria.

2.4. Data extraction and analysis

The full-text articles were obtained, and two researchers (K.K. and U. M.) independently reviewed all studies to determine eligibility for final inclusion in this review. Any disagreements were resolved by consensus. Any P4P programs that target surgical care and that are explicitly framed as reward, penalty, or combination of both in the United States were considered. The primary outcomes of interest were quality and cost

Table 1

Search terms used in hospital setting pay-for-performance for surgical care literature review.

Database/Searching Engine	Search Terms	Results
PubMed	(“pay for performance”[tiab] OR P4P[tiab] OR “pay for value”[tiab] OR “financial penalties” OR “financial incentive” OR ((bonus[tiab] OR reward[tiab] OR penalty [tiab] OR nonpayment[tiab]) AND (payment[tiab] OR reimburse*[tiab] OR incentive*[tiab] OR penalty*[tiab] OR nonpayment*[tiab]) AND (quality[tiab] OR value[tiab])) AND (hospital[tiab] OR inpatient[tiab]) AND (surgery[tiab] OR perioperative[tiab])	116
Embase	‘pay for performance’:ab,ti OR p4p:ab,ti OR ‘pay for value’:ab,ti OR ‘nonpayment’: ab,ti AND ‘hospital patient’ AND [2003–2020]/py	147
EconLit	(ab(‘pay for performance’) OR ab(p4p) OR ab(incentive) OR ab(penalty) OR ab (nonpayment)) AND ab(inpatient)	72
DARE	‘pay for performance’ and MeSH DESCRIPTOR Reimbursement, Incentive EXPLODE ALL TREES	21
CDSR	“pay for performance” in Title, Abstract, Keywords or “nonpayment” in Title, Abstract, Keywords or “incentive” in Title, Abstract, Keywords or “penalty” in Title, Abstract, Keywords and “inpatient” in Title, Abstract, Keywords, Publication Year from 2003 to 2020	73
Selected P4P researcher search (PubMed & Google Scholar)	Researcher list: Andrew Ryan, Adams Dudley, Howard Beckman, Kathleen Curtin, Larry Casalino, Tim Doran, Ashish Jha, Laura Petersen, Martin Roland, Meredith Rosenthal, Eric Schneider, Rachel Werner, Cheryl Damberg ●PubMed search term: (Last Name, First Name [Author]) ●Google Scholar search term: allintitle: ‘pay for performance’ OR ‘hospital’; author: “First name Last name”	2383
Other Sources	CMS, RAND, Commonwealth Fund, Kaiser Permanente, and Reference list review	31

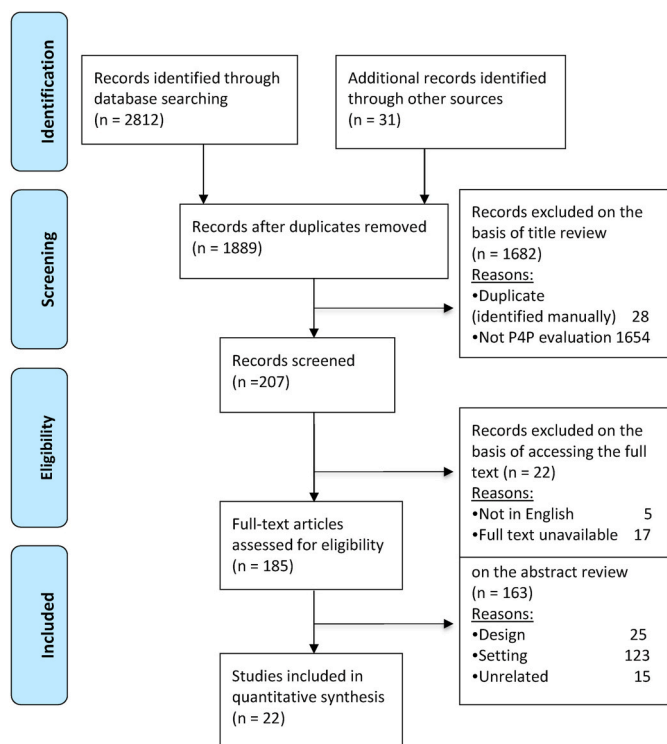


Fig. 1. PRISMA^a flow diagram for selected studies in the systematic literature review.

because these outcomes directly demonstrate the ultimate outcome that matters to patients and society [19]. Fig. 1 presents the study selection process based on the PRISMA flow diagram [20]. The quality of the included studies was evaluated using the modified Oxford Centre for Evidence-Based Medicine [21]. We assessed the risk of bias using the risk of bias in non-randomized studies of interventions tool with GRADE approach [22] and summarized this in Fig. 2.

3. Results

3.1. Characteristics of eligible studies

This review included 22 studies that met the inclusion criteria. The majority of the studies were published after 2010. The study population included inpatient surgical patients for the following procedures: coronary artery bypass graft [CABG] (ten studies), trauma (one study), bariatric surgery (two studies), orthopedic surgery (nine studies), vascular surgery (one study), obstetrics (one study), and general surgical care (five studies). Some studies evaluated more than one surgical procedure. Twelve studies included patients who underwent surgery for non-targeted conditions or patients in non-P4P program-participating hospitals as the control group. The other ten studies did not have a control group. Two studies used propensity score matching, eight a difference-in-differences (DD) or difference-in-difference-in-differences (DDD) design, two used the DD with a propensity score matched sample, four studies applied an interrupted time series analysis, and one used a regression discontinuity design. The average quality rating of the studies based on the Oxford Centre for Evidence-Based Medicine criteria was 3 (range 2–4), where level 2 studies represent those yielding the highest level of individual studies. We assessed that 12 (55%), 8 (36%), and 2 (9%) of the 22 studies had no serious, serious, and more than very serious risk of bias, respectively, mainly due to confounding factors and measurement of outcomes. (See S1 Table for an overview of studies analyzed, including sample, design, outcomes, results and other study characteristics).

Domains	confounding	Selection of participants	Classification of interventions	Deviations from intended interventions	Missing data	Measurement of outcomes	Selection of the reported result	GRADE RoB
Atkinson et al., 2010 [40]	Yellow	Green	Blue	Blue	Grey	Green	Blue	Red
Berthiaume et al., 2016 [33]	Red	Blue	Blue	Blue	Grey	Blue	Green	Red
Casale et al., 2007 [37]	Yellow	Blue	Blue	Blue	Grey	Blue	Blue	Orange
Epstein et al., 2014 [41]	Green	Blue	Blue	Blue	Blue	Blue	Blue	Dark Blue
Gu et al., 2014 [28]	Green	Green	Blue	Blue	Blue	Blue	Yellow	Orange
Healy et al., 2012 [24]	Green	Green	Blue	Blue	Blue	Blue	Blue	Orange
Ibrahim et al., 2017 [27]	Green	Blue	Blue	Blue	Blue	Green	Blue	Orange
Jha et al., 2012 [39]	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Dark Blue
Kandilov et al., 2014 [23]	Green	Blue	Blue	Blue	Blue	Green	Blue	Orange
Ryan et al., 2011 [31]	Blue	Blue	Blue	Blue	Grey	Green	Blue	Dark Blue
Ryan et al., 2015 [34]	Blue	Blue	Blue	Blue	Blue	Green	Blue	Dark Blue
Ryan et al., 2010 [35]	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Dark Blue
Ryan et al., 2009 [36]	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Dark Blue
Ryan et al., 2012 [42]	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Dark Blue
Scalea et al., 2014 [32]	Yellow	Green	Blue	Blue	Grey	Blue	Blue	Orange
Shih et al., 2014 [38]	Blue	Blue	Blue	Blue	Grey	Blue	Blue	Dark Blue
Hsu et al., 2018 [25]	Green	Blue	Blue	Blue	Blue	Blue	Blue	Orange
Thirukumar et al., 2019 [26]	Blue	Green	Blue	Blue	Blue	Blue	Blue	Dark Blue
Sankaran et al., 2019 [13]	Blue	Blue	Blue	Blue	Blue	Green	Blue	Dark Blue
Mehtsun et al., 2018 [29]	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Dark Blue
Li et al., 2019 [15]	Yellow	Green	Blue	Blue	Blue	Blue	Blue	Orange
Ramaswamy et al., 2019 [30]	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Dark Blue

Fig. 2. Risk of bias assessment of included studies using the risk of bias in non-randomized studies of interventions tool with GRADE approach.

3.2. Key features of the P4P programs

Key features of the P4P programs (i.e., program name, implementation year, target procedures, payment design, payment size, number of participating hospitals, and other important features) are presented in Table 2. Eleven specific P4P programs were included in this review. We classified P4P programs into three categories of payment design: penalties, rewards, and a combination of both. Three programs used penalties; five P4P programs used rewards; and three P4P programs used a combination of both. Six of 11 programs were initiated by CMS. The majority of targeted surgical procedures of P4P programs in the studies were CABG, hip fracture, THA, and TKA. Table 3 shows the

Table 2
Key Features of P4P programs Targeted Surgical Care in the Systematic Review.

P4P program	Ref#	Implement years & Other features	What to incentivize		Who to incentivize	How to incentivize				
			Outcome vs. Input	Specific vs.broad target conditions	Individual providers vs.hospital	Reward vs.penalty vs. withholding	Absolute vs.relative performance	Payment size	Voluntary vs. Mandatory	Numbers of participating hospitals
FIP	[32]	<ul style="list-style-type: none"> Feb 2013 – Unclear (Results reported through Dec 2013) A single hospital initiative 	Outcome	Trauma	Provider	Reward	Absolute	Certified registered nurse anesthetist: \$100–200/2weeks Registered nurse: \$50–120/2weeks Scrub technician: \$40–80/2weeks	Voluntary	1
HAC-POA	[23, 24, 25, 26]	<ul style="list-style-type: none"> Q4 2008 – Present CMS initiative Claim basis 	Outcome	SSI following CABG, orthopedics (spine, neck, shoulder, elbow), and bariatric surgery, DVT & PE following THA or TKA	Hospital	Penalty	Absolute	No payment for hospital acquired conditions	Mandatory	3203 (FY 2017)
HACRP	[13]	<ul style="list-style-type: none"> Q4 2014 – Present Present CMS initiative Claim basis 	Outcome	Hospital-associated infections including surgical site infection following abdominal hysterectomy and colon procedures	Hospital	Penalty	Relative	1% payment reduction	Mandatory	3306 (FY 2017)
HQSR	[33]	<ul style="list-style-type: none"> 2001 – Present Payer initiative 	Process & Outcome	Surgical and obstetrical procedures	Hospital	Reward	Relative	Varies by hospitals (reward size depends on hospital's share of the total Hawaii Medical Service Association payout (\$9 million in sum in 2004)	Voluntary	17
HRRP	[15,27–30]	<ul style="list-style-type: none"> Q4 2012 – Present CMS initiative Payment adjustment to all Medicare discharges 	Outcome	THA, TKA, CABG	Hospital	Penalty	Absolute	Up to 3% [Base DRG payment amount * readmissions adjustment factor] - Base operating DRG payment amount	Mandatory	3129 (FY 2020)
HVBP	[34]	<ul style="list-style-type: none"> Q4 2012 – Present CMS initiative Budget neutral (all withheld monies are paid out as rewards) 	Process & Outcome	Surgical cases	Hospital	Withholding & Incentive	Relative	2%	Mandatory	2955 (FY 2017)
Long Island Provider Initiated P4P	[40]	<ul style="list-style-type: none"> 2004 – Present Provider initiative 	Process & Outcome	Patient satisfaction, Patient safety, Hospital quality for conditions including, but are not limited to hip fracture, and other surgical procedures	Hospital	Reward & Penalty	Relative	50% of at-risk amount for each hospital	Voluntary	10
Mass Health	[31]	<ul style="list-style-type: none"> FY 2008 – Present MA State initiative Medicaid patients Greater incentive than PHQID 	Process & outcome (health disparity)	SSI	Hospital	Reward	Relative (1) Performance above median of all hospitals (2)Top decile to earn maximum (3)Improve-ment from previous year	\$25,000,000 (Rate year 2017) If divided equally among the 66 hospitals, this allocation would have exceeded \$370,000 per hospital	Voluntary	66
PHQID	Phase I [35,36]	<ul style="list-style-type: none"> Phase I Q4 2004 – 		CABG, THA, TKA	Hospital	Reward	Relative	Reward (1) 2% for the top decile	Voluntary	265

(continued on next page)

Table 2 (continued)

P4P program	Ref#	What to incentivize		Who to incentivize		How to incentivize		Payment size	Voluntary vs. Mandatory	Numbers of participating hospitals
		Implement years & Other features	Outcome vs. Input	Specific vs. broad target conditions	Individual providers vs. hospital	Individual providers vs. hospital	Reward vs. penalty vs. withholding			
PHQID		Q3 2006 • CMS initiative Phase II [38, 39, 41, 42]	• Phase II Q4 2006 – Q3 2009 • CMS initiative		CABG, THA, TKA	Hospital	Reward & Penalty	Relative	Voluntary	233
Proven care	[38]	• Feb 2006 – Present • Provider initiative	Process CABG, PCI, Cataract, THA, TKA, hip fracture, bariatric, and low back surgeries and perinatal		Provider	Reward	Absolute	100%	Voluntary	740 physicians at 3 hospitals and 40 primary care clinics covered by Geisinger Health Plan

summary characteristics of P4P programs included in this review, with substantial variability across programs. The majority of programs aimed at hospitals (n = 9, 82%). All Mandatory P4P programs were introduced by CMS. The first penalty-, reward-, and combination of reward and penalty programs were implemented in 2008, 2001, and 2004, respectively.

Note: Surgical procedures are in italics. AMI: acute myocardial infarction, CABG: coronary artery bypass graft, CAUTI: catheter-associated urinary tract infections, CDI: clostridium difficile infection, CHF: congestive heart failure, CLABSI: central line-associated bloodstream infections, DVT: deep vein thrombosis, CMS: Centers for Medicare and Medicaid Services, FIP: financial incentive program, HAC-POA: hospital acquired conditions presented on admission, FY: fiscal year, HACRP: hospital acquired condition reduction program, HQSR: hospital quality service and recognition, HRRP: hospital readmission reduction program, HVBP: hospital value-based purchasing program, MA: Massachusetts, MRSA: methicillin-resistant staphylococcus aureus, PCI: Percutaneous Coronary Intervention, PE: pulmonary embolism, PHQID: premier hospital quality incentive demonstration, PN: pneumonia, SSI: surgical site infection, THA: total hip arthroplasty, TKA: total knee arthroplasty, VCAI: vascular catheter-associated infections.

Effects of P4P programs on quality and cost of surgical care.

Twenty-one of the 22 studies evaluated quality outcomes associated with P4P programs using penalty (10), reward (7), or a combination of reward and penalty (4). Four of the 22 studies evaluated the cost of care related to P4P programs using penalty (1), reward (2), and the combination of both (1). Table 4 summarizes the number of studies by payment design of P4P programs (i.e., penalty, reward, or the combination of both) and the direction of effects (i.e., positive, negative, null, or mixed). The quality outcomes assessed in these P4P programs varied among studies and included mortality, readmissions, length of stay (LOS), number of procedures, number of complications, composite quality process score, operating room (OR) turnover time, and on-time start of the first surgery of the day. The cost outcomes evaluated in these programs include program operation costs, Medicare costs, and payments received by hospitals.

Multiple outcomes were evaluated in 12 out of 22 studies. Thus, the numbers may not sum to the total numbers of studies. ↑ indicates that the P4P improved the outcomes; ↓ indicates that the P4P worsened the outcomes; ∅ indicates a null effect; and ↑↓ indicates inconsistent findings (i.e., some outcomes improved with the P4P, whereas others showed no improvement, or improvement was also reported in the control group in the same study) or that no conclusion can be drawn. See S2 Table for the references.

3.3. Quality of surgical care for penalty-based programs

Ten studies examined outcomes of penalty based P4P programs on the quality of surgical care [13,15,23] [–] [30], with the majority showing that penalty designs were effective [24,26,27,29,30]. Although six out of 10 studies reported improved quality in surgical care, we conclude that one study did not congruently show a positive effect of P4P programs due to lack of comparison to the pre-policy period [23]. Two studies observed a positive spillover effect of a penalty based P4P program for nontargeted procedures [27] or non-Medicare patients [24].

3.4. Quality of surgical care for reward-based programs

The seven studies evaluating the effect of P4P with reward design had mixed results [31–37]. Two of the seven studies found a significant effect of P4P implementation on at least one of the outcomes; one found a positive effect [32], while the other study found a negative effect [35]. Positive effects included reduced turnover time in the OR and improved on-time first case starts while negative effects included reduced access to surgical care for non-White, Black, or Hispanic cardiac patients. Two of

Table 3
Summary Characteristics of P4P Programs Included in the Systematic Review. Rows may not add to 100% due to the programs targeted multiple surgical procedures.

Characteristics of P4P programs	No. of P4P programs/Total No. of P4P programs (%)
Payment Strategy	
Penalties ^a	3/11 (27%)
Rewards ^b	5/11 (45%)
Combination of rewards and penalties ^c	3/11 (27%)
Targeted Surgical Procedures	
All surgical procedures	5/11 (45%)
Cardiac procedures ^d	4/11 (36%)
Orthopedic procedures ^e	4/11 (36%)
Others ^f	2/11 (18%)
Targets	
Providers	2/11 (18%)
Hospitals	9/11 (82%)
Initiator	
Payer ^g	7/11 (63%)
Hospital/Provider	3/11 (27%)
State	1/11 (9%)
Earliest Program Implementation years	
Penalty	2008
Reward	2001
Combination of reward and penalty	2004
	Median (Interquartile range)
Participating Hospitals	233 (10–3129)

^a Penalty-based programs included in this review are: hospital acquired conditions presented on admission, hospital acquired condition reduction program, and hospital readmission reduction program.

^b Reward-based programs are: financial incentive program, hospital quality service and recognition, MassHealth, the first phase of the premier hospital quality incentive demonstration (PHQID), and ProvenCare

^c Combination of rewards and penalty programs are: hospital value-based purchasing program, Long Island provider-initiated P4P, and the second phase of the PHQID.

^d Cardiac procedures include coronary artery bypass graft and percutaneous transluminal coronary angioplasty.

^e Orthopedic procedures include spine, shoulder, elbow, knee, hip, and trauma surgeries.

^f Others include trauma, vascular, and obstetrics procedures.

^g Payers include the Centers for Medicare & Medicaid Services and a private health insure.

Table 4
Summary of the effects of P4P programs included in the systematic review.

Measurement of Effects of P4P programs	Payment design											
	Penalty				Reward				Combination of Reward and Penalty			
	Direction of Effect				Direction of Effect				Direction of Effect			
	↑	↓	∅	↑↓	↑	↓	∅	↑↓	↑	↓	∅	↑↓
	No. of studies				No. of studies				No. of studies			
Quality of Surgical Care												
Composite quality process score							2					1
Number of complications/Complication rates	2		2					1			1	
Length of stay	1							2				1
Mortality ^a			1				1				2	
Readmission ^b	3		2	1								
Others ^c	2											1
Cost of Surgical Care ^d				1	1	1	1	1	1		1	

^a Mortality includes 30-day mortality and operative mortality.

^b Readmission include 30-day readmission, 90-day readmission, and readmission to ICU.

^c Others include operating room turnover time, operating room on time 1st case start, patient experience, access to care (i.e., number of patients who have undergone CABG procedure and the rate at which patients undergo CABG procedure), use of skilled nursing facilities, use of blood products, reintubation during hospital stay, total ventilation hours, percentage of hospitals that received incentives, and spillover effect to nontargeted procedures or other payers. These were grouped because only one study examined these outcomes.

^d The measures of surgical care cost include Medicare payment, hospital payment, physician payment, post-acute care payment, home health agency payment, the program's potential savings, and payment received across incentivized conditions.

the seven studies reported improvements related to P4P program implementation [33,37], but the results were based on the pre-post comparison or lacked details on statistical methods, and consequently no conclusion can be drawn.

3.5. Quality of surgical care for combined reward and penalty-based programs

The four studies that evaluated the effects of P4P with combined payment design did not show any significant results [38–41]. Three of the studies that evaluated the PHQID, using reward payment in the first phase and a combined reward and penalty payment for the second phase, did not observe a significant improvement or a negative impact on patient outcomes [38,39,41]. The fourth study found a decreased LOS and an improved quality of care [40]. However, this study was descriptive, and no statistical significance was reported.

3.6. Cost of surgical care

Four of the 22 studies investigated the effect of P4P on the cost of surgical care using a penalty [23], reward [32,36], and the combination of reward and penalty [42]. One study evaluated a single hospital-initiated P4P program with a reward for providers and reported improved efficiency in the OR and estimated savings [32]. Three studies that examined CMS-initiated P4P programs reported mixed effects [23, 36,42]. One of the three studies investigated P4P with rewards and found no evidence for a reduction in costs for CABG patients [36]. Another study consisted of a descriptive analysis of surgical care costs after implementation of the CMS P4P that utilizes a penalty design and was not a comparison of the pre- and post-implementation [23]. Therefore, this study did not provide evidence for improvement in the costs of care related to the P4P program. The other study compared the effect of a P4P payment program that used a combined reward and penalty design with different incentive sizes based on quality outcomes post implementation [42]. They found that hospitals with larger shares of socioeconomically disadvantaged patients received more incentives. However, this was due to an incentive size change that was introduced towards the later phase of the program, not due to a quality improvement. We conclude from these studies that evidence regarding the effects of P4P programs on cost is limited and remains largely unknown.

4. Discussion

This systematic review evaluated the effects of different P4P payment designs. We found mixed evidence on the effectiveness of P4P in quality improvement. However, with five out of 10 studies reporting positive effects of penalty-based programs, and studies evaluating P4P programs with a reward design or a combination of rewards and penalties showing little or null effects, penalty-based payment design emerged as a potentially more effective strategy than reward-based programs or programs using rewards and penalties in combination.

One explanation for why penalty-based P4P programs might be more effective than other designs is the tendency for people and organizations to respond to losses more than to gains, a behavior explained by the theory of loss aversion [43]. Supported by behavioral economics and psychology, this theory postulates that penalties may induce stronger provider and hospital motivation to improve quality. Subjecting hospitals to payment reductions on TKA, THA, and CABG alone, which are the most frequently targeted surgical procedures by P4P programs with over 1.5 million procedures performed annually [2], likely incentivizes change in behavior.

We found inter- and intra-program variations. Regarding inter-program variations within the same penalty-based payment design, the hospital readmission reduction program was shown to be more likely to have positive impacts on improving quality of care. This might be related to the penalty size. The hospital readmission reduction program imposes up to 3% penalties for each patient if hospitals have a higher number of 30-day readmissions for the targeted conditions, which has impacted 83% hospitals [44]. Intra-program variations were also found. Whereas some studies evaluated effects over at least 5 years and showed positive effects of the hospital readmission reduction program on decreasing readmissions, others evaluated the effects over a short period and found no significant effect. This suggests the importance of long-term program evaluation because it could take years for hospitals to change their practices and to observe improvements in quality [39].

Our review shows the small number of studies that have evaluated the impact of P4P on cost outcomes for surgical care, although one of the main purposes of P4P is cost reduction. One reason for the lack of studies focusing on cost outcomes could be an uneasiness in the United States to ration health care based on cost [45]. Some policymakers argue against using cost-effectiveness as a reason to cut expensive yet effective treatments [46]. The Patient-Centered Outcomes Research Institute [47] also prohibits the use of a dollar-per-quality-outcome measurement in establishing treatment recommendations [48]. Despite the ethical complexities of health care rationing that underline these recommendations, payers and many other policymakers have had a substantial interest in implementing P4P programs to curb vast health expenditures in resource-scarce settings [49].

With the continued growth of P4P programs utilizing penalties, the number of studies evaluating penalty-based P4P programs have increased. We did not find randomized controlled trials (RCTs), as RCTs are rarely used to evaluate P4P programs due to the high costs of RCTs [50]. However, recent studies have used advanced analytic methods, such as a DD or DDD model, interrupted time series, and propensity score matching, to overcome the limitations of observational data. These studies have produced robust results regarding the effects of P4P, which allowed us to conclude that penalty designs could improve the quality of surgical care more effectively than other payment designs. Future studies with methodologically rigorous designs are needed to verify the potential of penalty-based P4P programs over other types of payment design programs.

This review has three notable limitations. First, a meta-analysis could not be conducted because of the lack of studies evaluating the same outcomes. If future studies become available, a meta-analysis may be able to produce quantifiable summary data on the effectiveness of specific payment designs. Second, studies examined P4P programs with substantially varied design features, as described. The aim of this review

was to evaluate the variation in quality and cost outcomes of surgical care as attributable to P4P payment designs, thus we compared findings across studies solely focused on the type of payment strategy and did not analyze heterogeneity across other features, such as incentive size and mandatory versus voluntary participation. Third, our findings may not be generalizable to all surgical procedures because the majority of studies examined P4P programs that targeted CABG, THA, or TKA only.

5. Conclusion

Results of this systematic review suggest that P4P programs utilizing penalties could be more effective than those utilizing rewards or a combination of both to improve the quality of surgical care. P4P has been used as a tool to improve quality at a reduced cost since the early 2000s and will likely function as an important policy tool for the foreseeable future. With the growing volume and high costs of surgical procedures, P4P initiatives are increasingly popular in surgical settings. Considering that over 4400 surgeries per 100,000 people are performed annually at acute care hospitals [2], the implications of improving surgical care quality are significant.

Funding/support

The research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Provenance and peer review

Not commissioned, externally peer-reviewed.

Ethical approval

This is a review article and received an exemption of review.

Consent

n/a

Author contribution

K Kim: conceptualization, study design, data collection, data analysis, writing – original draft and review & editing.

W Max: conceptualization, study design, writing – review and editing.

J White: conceptualization, study design, writing – review and editing.

S Chapman: conceptualization, study design, writing – review and editing.

U Muench: conceptualization, study design, data analysis, writing – review and editing.

Registration of Research Studies

Name of the registry: Research Registry

Unique Identifying number or registration ID: reviewregistry944

Hyperlink to your specific registration (must be publicly accessible and will be checked): <https://www.researchregistry.com/browse-the-registry/#registryofsystematicreviewsmeta-analyses/registryofsystematicreviewsmeta-analysesdetails/5effa270994afe001534394b/>

Guarantor

K Kim

U Muench

Declaration of competing interest

The authors have no conflicts of interest to disclose.

Acknowledgements

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.amsu.2020.11.060>.

References

- [1] E. Muñoz, W. Muñoz III, L. Wise, National and surgical health care expenditures, 2005–2025: *Ann. Surg.* 251 (2010) 195–200, <https://doi.org/10.1097/SLA.0b013e3181cbcc9a>.
- [2] K.W. McDermott, W.J. Freeman, A. Elixhauser, *Overview of Operating Room Procedures during Inpatient Stays in U.S. Hospitals, 2014*, Agency for Healthcare Research and Quality, 2017.
- [3] C.A. Steiner, Z. Karaca, B.J. Moore, M. Imshaug, G. Pickens, *Surgeries in Hospital-Based Ambulatory Surgery and Hospital Inpatient Settings, 2014, 2017*.
- [4] M.A. Healy, A.J. Mullard, D.A. Campbell, J.B. Dimick, Hospital and payer costs associated with surgical complications, *JAMA Surg.* 151 (2016) 823, <https://doi.org/10.1001/jamasurg.2016.0773>.
- [5] R. Busse, Pay-for-performance: time to act but also to provide further evidence, *Health Pol.* 120 (2016) 1123–1124, <https://doi.org/10.1016/j.healthpol.2016.10.001>.
- [6] R.A. Dudley, Pay-for-performance research: how to learn what clinicians and policy makers need to know, *J. Am. Med. Assoc.* 294 (2005) 1821–1823.
- [7] R. Milstein, J. Schreyoegg, Pay for performance in the inpatient sector: a review of 34 P4P programs in 14 OECD countries, *Health Pol.* 120 (2016) 1125–1140, <https://doi.org/10.1016/j.healthpol.2016.08.009>.
- [8] K.K. Kondo, C.L. Damberg, A. Mendelson, M. Motu'apuaka, M. Freeman, M. O'Neil, et al., Implementation processes and pay for performance in healthcare: a systematic review, *J. Gen. Intern. Med.* 31 (2016) 61–69, <https://doi.org/10.1007/s11606-015-3567-0>.
- [9] A.M. Ryan, J. Blustein, Making the best of hospital pay for performance, *N. Engl. J. Med.* 366 (2012) 1557–1559, <https://doi.org/10.1056/NEJMp1202563>.
- [10] S.R. Kristensen, Financial penalties for performance in health care, *Health Econ.* 26 (2017) 143–148, <https://doi.org/10.1002/hec.3463>.
- [11] M. Abdellaoui, H. Bleichrodt, C. Paraschiv, Loss aversion under prospect theory: a parameter-free measurement, *Manag. Sci.* 53 (2007) 1659–1674, <https://doi.org/10.1287/mnsc.1070.0711>.
- [12] T. Rouyard, A. Attema, R. Baskerville, J. Leal, A. Gray, Risk attitudes of people with 'manageable' chronic disease: an analysis under prospect theory, *Soc. Sci. Med.* 214 (2018) 144–153, <https://doi.org/10.1016/j.socscimed.2018.08.007>.
- [13] R. Sankaran, D. Sukul, U. Nuliyalu, B. Gulseren, T.A. Engler, E. Arntson, et al., Changes in hospital safety following penalties in the US Hospital Acquired Condition Reduction Program: retrospective cohort study, *BMJ* (2019), l4109, <https://doi.org/10.1136/bmj.l4109>.
- [14] A. Mendelson, K. Kondo, C. Damberg, A. Low, M. Motu'apuaka, M. Freeman, et al., The effects of pay-for-performance programs on health, health care use, and processes of care: a systematic review, *Ann. Intern. Med.* 166 (2017) 341, <https://doi.org/10.7326/M16-1881>.
- [15] B.Y. Li, K.L. Urish, B.L. Jacobs, C. He, T. Borza, Y. Qin, et al., Inaugural readmission penalties for total hip and total knee arthroplasty procedures under the hospital readmissions reduction program, *JAMA Netw. Open* 2 (2019), e1916008, <https://doi.org/10.1001/jamanetworkopen.2019.16008>.
- [16] 2017 report Health Care Cost and Utilization Report. Health Care Cost Institute; 2019.
- [17] M.A. Healey, S.R. Shackford, T.M. Osler, F.B. Rogers, E. Burns, Complications in surgical patients, *Arch. Surg.* 137 (2002) 8.
- [18] C.L. Damberg, M.E. Sorbero, S.L. Lovejoy, G.R. Martzolf, L. Raaen, D. Mandel, Measuring Success in Health Care Value-Based Purchasing Programs: Findings from an Environmental Scan, Literature Review, and Expert Panel Discussions, RAND Corporation, Santa Monica, CA, 2014.
- [19] M.E. Porter, S. Larsson, T.H. Lee, Standardizing patient outcomes measurement, *N. Engl. J. Med.* 374 (2016) 504–506.
- [20] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, *Ann. Intern. Med.* 151 (2009) 264–269.
- [21] OCEBM Levels of Evidence Working Group. Oxford 2011 levels of evidence n.d.
- [22] H.J. Schünemann, C. Cuello, E.A. Akl, R.A. Mustafa, J.J. Meerpohl, K. Thayer, et al., GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence, *J. Clin. Epidemiol.* 111 (2019) 105–114, <https://doi.org/10.1016/j.jclinepi.2018.01.012>.
- [23] A. Kandilov, N. Coomer, K. Dalton, The impact of hospital-acquired conditions on Medicare program payments, *Medicare Medicaid Res. Rev.* 4 (2014) E1–E23, <https://doi.org/10.5600/mmrr.004.04.a01>.
- [24] D. Healy, J. Cromwell, *Hospital-acquired Conditions – Present on Admission: Examination of Spillover Effects and Unintended Consequences*, Centers for Medicare and Medicaid Services, 2012.
- [25] H.E. Hsu, A.T. Kawai, R. Wang, M.S. Jentzsch, C. Rhee, K. Horan, et al., The impact of the Medicaid healthcare-associated condition program on mediastinitis following coronary artery bypass graft, *Infect. Control Hosp. Epidemiol.* 39 (2018) 694–700, <https://doi.org/10.1017/ice.2018.69>.
- [26] C.P. Thirukumar, L.G. Glance, M.B. Rosenthal, H. Temkin-Greener, R. Balkissoon, A. Mesfin, et al., Impact of medicare's nonpayment program on venous thromboembolism following hip and knee replacements, *Health Serv. Res.* 53 (2018) 4381–4402, <https://doi.org/10.1111/1475-6773.13013>.
- [27] A.M. Ibrahim, H. Nathan, J.R. Thumma, J.B. Dimick, Impact of the hospital readmission reduction program on surgical readmissions among medicare beneficiaries, *Ann. Surg.* 266 (2017) 617–624, <https://doi.org/10.1097/SLA.0000000000002368>.
- [28] Q. Gu, L. Koenig, J. Faerber, C.R. Steinberg, C. Vaz, M.P. Wheatley, The Medicare Hospital Readmissions Reduction Program: potential unintended consequences for hospitals serving vulnerable populations, *Health Serv. Res.* 49 (2014) 818–837, <https://doi.org/10.1111/1475-6773.12150>.
- [29] W.T. Mehtsun, I. Papanicolas, J. Zheng, E.J. Orav, K.D. Lillemo, A.K. Jha, National trends in readmission following inpatient surgery in the hospital readmissions reduction program era, *Ann. Surg.* 267 (2018) 599–605, <https://doi.org/10.1097/SLA.0000000000002350>.
- [30] A. Ramaswamy, M. Marchese, A.P. Cole, S. Harmouch, D. Friedlander, J. S. Weissman, et al., Comparison of hospital readmission after total hip and total knee arthroplasty vs spinal surgery after implementation of the hospital readmissions reduction program, *JAMA Netw. Open* 2 (2019), e194634, <https://doi.org/10.1001/jamanetworkopen.2019.4634>.
- [31] A.M. Ryan, J. Blustein, The effect of the MassHealth hospital pay-for-performance program on quality, *Health Serv. Res.* 46 (2011) 712–728, <https://doi.org/10.1111/j.1475-6773.2010.01224.x>.
- [32] T.M. Scalea, D. Carco, M. Reece, Y.L. Fouche, A.N. Pollak, S.S. Nagarkatti, Effect of a novel financial incentive program on operating room efficiency, *JAMA Surg.* 149 (2014) 920–924, <https://doi.org/10.1001/jamasurg.2014.1233>.
- [33] J.T. Berthiaume, R.S. Chung, K.L. Ryskina, J. Walsh, A.P. Legorreta, Aligning financial incentives with quality of care in the hospital setting, *J. Health C Qual.* 28 (36–44) (2006) 51.
- [34] A.M. Ryan, J.F.J. Burgess, M.F. Pesko, W.B. Borden, J.B. Dimick, The early effects of Medicare's mandatory hospital pay-for-performance program, *Health Serv. Res.* 50 (2015) 81–97, <https://doi.org/10.1111/1475-6773.12206>.
- [35] A.M. Ryan, Has pay-for-performance decreased access for minority patients? *Health Serv. Res.* 45 (2010) 6–23, <https://doi.org/10.1111/j.1475-6773.2009.01050.x>.
- [36] A.M. Ryan, Effects of the premier hospital quality incentive demonstration on Medicare patient mortality and cost, *Health Serv. Res.* 44 (2009) 821–842, <https://doi.org/10.1111/j.1475-6773.2009.00956.x>.
- [37] A.S. Casale, R.A. Paulus, M.J. Selna, M.C. Doll, A.E.J. Bothe, K.E. McKinley, et al., "ProvenCareSM": a provider-driven pay-for-performance program for acute episodic cardiac surgical care, *Ann. Surg.* 246 (2007) 613–621, <https://doi.org/10.1097/SLA.0b013e318155a996>, discussion 621–623.
- [38] T. Shih, L.H. Nicholas, J.R. Thumma, J.D. Birkmeyer, J.B. Dimick, Does pay-for-performance improve surgical outcomes? An evaluation of phase 2 of the premier hospital quality incentive demonstration, *Ann. Surg.* 259 (2014) 677–681, <https://doi.org/10.1097/SLA.0000000000000425>.
- [39] A.K. Jha, K.E. Joynt, E.J. Orav, A.M. Epstein, The long-term effect of premier pay for performance on patient outcomes, *N. Engl. J. Med.* 366 (2012) 1606–1615.
- [40] J.G. Atkinson, K.E. Masulis, L. Felgner, D.N. Schumacher, Provider-initiated pay-for-performance in a clinically integrated hospital network, *J. Health C Qual.* 32 (2010) 42–50, <https://doi.org/10.1111/j.1945-1474.2009.00063.x>, quiz 50.
- [41] A.M. Epstein, A.K. Jha, E.J. Orav, The impact of pay-for-performance on quality of care for minority patients, *Am. J. Manag. Care* 20 (2014) e479–486.
- [42] A.M. Ryan, J. Blustein, T. Doran, M.D. Michelow, L.P. Casalino, The effect of Phase 2 of the Premier Hospital Quality Incentive Demonstration on incentive payments to hospitals caring for disadvantaged patients, *Health Serv. Res.* 47 (2012) 1418–1436, <https://doi.org/10.1111/j.1475-6773.2012.01393.x>.
- [43] D. Kahneman, A. Tversky, Choices, values, and frames, *Am. Psychol.* 39 (1984) 341–350.
- [44] J. Rau, New Round of Medicare Readmission Penalties Hits 2,583 Hospitals, Kaiser Health News, 2019 accessed, <https://khn.org/news/hospital-readmission-penalties-medicare-2583-hospitals/>. (Accessed 11 November 2020).
- [45] J.J. Kim, The role of cost-effectiveness in US vaccination policy, *N. Engl. J. Med.* 365 (2011) 1760–1761.
- [46] M.C. Weinstein, J.A. Skinner, Comparative effectiveness and health care spending-implications for reform, *N. Engl. J. Med.* 362 (2010) 460–465, <https://doi.org/10.1056/NEJMsb0911104>.
- [47] Patient-Centered Outcomes Research Institute, accessed, <https://www.pcori.org/>, 2019. (Accessed 20 February 2019).
- [48] O. Social Security Administration, Limitations on Certain Uses of Comparative Clinical Effectiveness Research, 2018 accessed, https://www.ssa.gov/O_P_Home/ssact/title11/1182.htm. (Accessed 12 September 2018).
- [49] J. James, *Health Policy Brief: Pay-For-Performance*, Health Affairs, 2012.
- [50] K. Mullen, R. Frank, M. Rosenthal, Can you get what you pay for? Pay-for-performance and the quality of healthcare providers, *Rand J. Econ.* vol. 41 (2010) 64–91.