Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

hche | Hamburg Center for Health Economics

A review of pay-for-performance programs in the inpatient sector in OECD countries

Ricarda Milstein Jonas Schreyögg

Research Paper
Year: 2015
No. 9

A review of pay-for-performance programs in the inpatient sector in OECD countries

Ricarda Milstein Jonas Schreyögg

hche Research Paper No. 09 (2015)
http://www.hche.de

Abstract

Background: Across the member countries of the Organisation for Economic Co-operation and Development (OECD), pay-for-performance (P4P) programs have been implemented in the inpatient sector to improve the quality of care provided by hospitals. However, little is known about whether such programs can live up to expectations. Thus far, evaluations and reviews have focused on the ambulatory care sector in Anglo-Saxon countries. The transferability of lessons learned to the inpatient sector, however, is limited.

Objectives: We aimed to provide an overview of existing P4P programs in the inpatient sector in the OECD countries and to assemble information on their effects. Furthermore, we attempted to identify whether evaluations of such programs allow preliminary conclusions to be drawn about the effects of P4P.

Methods: We conducted a structured literature search in five databases to identify relevant sources in Danish, English, French, German, Hebrew, Italian, Japanese, Korean, Norwegian, Spanish, Swedish and Turkish. This was complemented by desk-based research. In selected cases, we contacted experts to validate our results and to add further information. Our research was restricted to the inpatient sector in OECD countries.

Results: We identified 30 P4P programs in 14 OECD countries. The programs were very heterogeneous in their design. First, they catered to different aims. Some programs followed a narrow approach and focused on improving the quality of care for a single medical condition, whereas others aimed at improving the quality of inpatient care more broadly. Second, the programs blended structural, process and outcome measures that targeted different stages of inpatient care pathways. Third, the financial rewards were

designed in various ways. Programs based their rewards either on an absolute or a relative score. Incentives included payment withholds, penalties, bonuses, or a combination thereof. The size of the incentive often amounted to approximately 0.1% of a hospital's budget or less, and never exceeded 4%. Lastly, the results of published evaluations of the P4P programs ranged from no effect to moderately positive effects. In cases where evaluations had positive results, the effect was seldom sustained and the causalities were unclear.

Conclusion: The results of our review indicate that P4P has been widely adopted across the OECD and become an integral part of the inpatient sector. The programs are very heterogeneous. The impact of P4P is unclear, and it may be that the moderately positive effects seen for some programs can be attributed to side effects, such as public reporting and increased awareness of data recording. Policy makers must decide whether the potential benefits of introducing a P4P program outweigh the potential risks within their particular national or regional context, and should be aware that P4P programs have yet not lived up to expectations.

Keywords: Hospitals; pay-for-performance; quality; health system; health policy
JEL Classification: I10, I11, I18.

Ricarda Milstein
Institute of Management in Health Care,
Chair of Business Administration, and
Hamburg Center for Health Economics

Universität Hamburg
Esplanade 36
20354 Hamburg
Germany

ricarda.milstein@wiso.uni-hamburg.de

Jonas Schreyögg
Institute of Management in Health Care,
Chair of Business Administration, and
Hamburg Center for Health Economics

Universität Hamburg
Esplanade 36
20354 Hamburg
Germany

Jonas.schreyoegg@wiso.uni-hamburg.de

We thank Matthew Gaskins for his helpful comments and language corrections, and Andrea Bükow for administrative assistance with this paper.

1. Introduction

Across the member countries of the Organisation for Economic Co-operation and Development (OECD), policy makers are concerned about improving the quality of care in their health systems. Virtually all have adopted, are in the process of adopting, or are discussing pay-for-performance (P4P) programs as a potential means to this end. In its narrow sense, P4P has been defined as a way to improve the quality of care through financial incentives [1]. It is based on the understanding that health providers can be extrinsically motivated by financial incentives – that is, if improvements in the quality of care are financially rewarded, greater efforts will be made to achieve better quality [2, 3]. Compared to other sectors, the inpatient sector is a late mover. In non-health sectors, such as education and public administration, programs were already implemented in the late 1960s [4]. Programs within the field of health emerged in the 1970s in the United States when private provider groups started to experiment with programs which targeted patients to improve their lifestyle and compliance with treatment paths, or providers to improve their activity and compliance with guidelines. Public programs which operate on a larger or nationwide scale followed in the late 1990s and took off in ambulatory care [5, 6]. Prominent examples include programs such as the Quality and Outcomes Framework for Primary Care in England, which was launched in 2004, the Practice Improvement Program in Australia, which was initiated in 1998, and the P4P programs in the sickness funds Clalit and Maccabi which started in 1998 and 2001 respectively [7, 8]. Programs in the inpatient sector followed closely after. A decisive point was the introduction of the Premier Hospital Quality Incentive Demonstration (HQID) representing the first public P4P program in the public hospital sector of the US [9]. This has served as a model for following P4P programs in England, South Korea and the United States and has motivated other countries to pilot P4P programs, as well. As of 2015, there were at least 30 programs in the inpatient sector in place. Several further countries discuss adopting P4P in the inpatient sector. Germany plans to introduce P4P in the inpatient sector in 2016 and Belgium has decided to pilot P4P from the second half of 2016 on [10, 11].

Despite the increasing uptake in P4P, it is not clear how successful this instrument is. The research undertaken on this subject to date has focused predominantly on programs that target office-based providers [7, 12, 13]. Drawing lessons from this research for the inpatient sector would be problematic. For example, hospitals are complex organizations, and their incentive structure differs markedly from that of individual physicians [14]. Furthermore, inpatient and office-based ambulatory care providers are generally financed through different systems. Across the OECD, hospitals are paid through diagnosis-related groups (DRGs), global budgets, or a combination of these. In contrast, office-based physicians are remunerated through a blend of capitation, fee-for-service, P4P, global budgets and other methods [15]. Each payment system has a strong effect the behavior of health providers, including how they deliver care [16, 17].

Additionally, the research conducted on P4P programs to date has often been limited to the experience of Anglo-Saxon countries [7], overlooking the increasing number of P4P programs located elsewhere. Particularly from a policy point of view, countries with a statutory health insurance (SHI) system may learn more from the experiences of countries with similar systems rather than those with a National Health Service (NHS) like that in England or the highly complex and varied system of private insurance and government programs in the US. Moreover, when evidence is drawn from countries with similar systems, there may be less need to invest in system adjustments [18, 19].

In this paper, we aim to shed light on the impact of P4P programs in the inpatient sector both within and beyond the Anglo-Saxon countries. First, we provide an overview of existing P4P programs. We then take a closer look at their aims, indicators, financial rewards and effects. Subsequently, we draw upon these findings to discuss the strengths and weaknesses of P4P programs. Lastly, we conclude by suggesting several points that policy makers might want to consider when designing a P4P system.

2. Methods

We conducted a search of the following five databases: Medline, the Cochrane Database of Systematic Reviews, the Centre for Reviews and Dissemination (CRD) database, the Database of Abstracts of Reviews of Effects (DARE), and the NHS Economic Evaluation Database (NHS EED). The following search terms were used: pay-for-performance, payment by results, payment for quality, performance payment, value-incentive payment, value-based purchasing, and financial incentive. To allow for the inclusion of programs that were no longer ongoing, we did not restrict our search by publication date. We did, however, limit our search to studies of programs implemented in OECD member states and the inpatient sector. Documents written in Danish, English, French, German, Hebrew, Italian, Japanese, Korean, Norwegian, Spanish, Swedish and Turkish were considered eligible. Our search was complemented by desk-based research of documents from ministries, other government agencies, and statutory bodies at the federal, state and regional levels. In selected cases, we contacted experts in the respective countries to validate our results and add further information. We classified the programs identified in our research into two groups based on whether they awarded bonuses or levied penalties for the quality of care provided for each medical condition separately or for a group of medical conditions.

Moreover, we classified the indicators used in each program as being structural, procedural or outcome-based as defined by Donabedian [20]. With this overall approach, we identified 30 different programs in 14 of the 36 OECD countries. Our research revealed a high degree of heterogeneity among the P4P programs. This likely reflects diverging opinions about how best to design P4P programs, as well as differing aims of their implementation. Our results are reported in the table below and discussed in greater detail in the following sections.

3. Results: Overview of P4P programs in the inpatient sector of OECD countries

	Name	Region	Date	No. of hospital	Aim(s)	No. of indication	Indi-cators			Reward			% of hospital budget	Effects	Side effects
							Structure	Process	Outcome	Bonus	Malus	Withhold			
AUS	Clinical Practice Improvement Payment	Queensland	2007 (ceased)	22/111	Improve quality of care for selected medical conditions	5		X		X			<1%	Unknown	Unknown
AUS	National emergency access targets, national elective surgery target	Nationwide	2012	289/244	Reduce waiting times	All ED / elective cases		X		X			Unknown.	Little improvement for NEAT, almost no improvement for NEST	Unknown
CAN	ED P4P	British Columbia	2007	14	Reduce waiting times	All cases in ED		X		X			< 1%	Improvement, but below government targets	
CAN	ED Wait Time Strategy	Ontario	2008	74	Reduce waiting times	All cases in ED		X		X			< 1%	Improvement, but below government targets	
CAN	Performance-based compensation	Ontario	2011	unknown	Improve adherence to national quality guidelines	Irrespective		X		X			Up to 20% of CEO's income	Unknown	Unknown
DEN	Journalauditindikatoren	Southern Denmark	2009	4	Increase share of patients with case managers	all inpatient cases		X		X	X		<1%	Increased share of patients with case managers	Contested whether patients have actually seen their case managers. Danger of "ticking the box"
ENG	Advancing Quality	Northwest England	2008	24	Reduce mortality, reduce costs, reduce length of stay	10		X				X	<4%	Reduced mortality and length of stay, improvement in process indicators, but no sustainable effect in the long-run	Unknown
ENG	Commissioning for Quality and Innovation	Throughout England	2009	Unknown	Depends on local agreements as targets are defined by local Clinical Commissioning Groups.	Depends on agreement	X	X	X	X	X	X	Up to 2.5% of the CCG's annual allocation for non-recurrent expenditure	Potentially improvement in process indicators	Very heterogeneous programs, high administrative burden

	Best practice tariffs	Throughout England	2010	Depends on indication	Improve adherence to guidelines, perform surgeries as day cases	65		X	X	X		X	<1%	Improvement in process indicators, effect on outcomes not clear	Administrative difficulties, but presumably lower burden than CQUIN
FRA	Indication financière à l'amélioration de la qualité (IFAQ 1)	Nationwide	2014	222	Predominantly to improve documentation	-		X		X			<0.5%	Unknown	Unknown
	Indication financière à l'amélioration de la qualité (IFAQ 2)	Nationwide	2016	490				X		X			<0.5%	Unknown	Unknown
ISR	ביצועים לפי תשלום (Pay for performance)	Tel Aviv	2009 (ceased in 2011)	1	Improve outcomes of heart and thorax surgeries, reduce costs of complications	2		X	X		X		<1%	Positive effect on process indicators, no effect on surgery outcomes	Unknown
	יקרו בל אירועי (Never-events)	Nationwide	2011	irrespective	Reduce occurrence of never-events	6		X			X		<1%	Unknown	Unknown
ITA	Applicazione del percorso assistenziale nei pazienti ultrasessantacinquenni con fratture di femore (PAFF)	Lazio	2009	-	Reduce waiting times for patients with hip replacements aged 65 and above	1		X			X		<1%	Decrease in waiting time for hip replacements, but heterogeneous across hospitals	Effects on waiting times in other areas unknown
	CEO remuneration	Tuscany	2006	-	Improve various aspects of care quality, such as waiting times for hip replacements in patients aged 65 and above	Unknown		X	X	X			Up to 20% of CEO's income	Has reduced waiting times, but improvement was very heterogeneous	Effects on waiting times in other areas unknown
JAP	質に基づく支払い, 回復期リハビリテーション病棟入院料において	Nationwide	2008	-	Improve health outcomes of stroke patients	1		X	X	X			<1%	Positive effect on process indicators, no effect on outcome indicators	Potential increase in preferred selection of good risks ("cherry-picking" of patients)
LUX	Incitants qualité	Nationwide	1998	-	Improve overall quality of care	13	X	X	X			X	2%	Unknown	Unknown
NOR	Kvalitetsbasert finansiering	4 regions	2014	-	Improve overall quality of care	-		X	X	X			0.5%	Unknown	Unknown
SWE	Målrelaterat ersättning	Stockholms läns landstin	2004	6	Improve adherence to national quality guidelines	Irrespective		X				X	2%	Unknown	Unknown
	Målrelaterat ersättning	Vårdgivare i Skåne	?	10	Improve adherence to national quality guidelines	respective		X				X	To follow	Unknown	Unknown

	Målrelaterat ersättning	Uppsala län	2005	2	Improve adherence to national quality guidelines	irrespective		X				X	4%	Unknown	Unknown
	Målrelaterat ersättning	Västra Götalands region	2005	18	Improve adherence to national quality guidelines	irrespective		X	X			X	3%	Unknown	Unknown
ROK	한국의가감지급시범사업 (Value incentive program)	Nationwide	2007	44	Reduce variation in quality of care, reduce fatality rates	4		X	X	X	X		2%	Improvement in average quality of care, reduced variation in use of procedures and patient outcomes among hospitals	Likely prone to unintended consequences
TUR	Performansa dayali ek ödeme sistemi (performance-based supplementary payment system	Nationwide	2004	-	Increase activity level, improve job satisfaction of physicians	irrespective	X	X	X	X			Up to 40% of physician's income	Has increased activity level	Growth in activity beyond medical appropriateness of care
USA	Hospital Quality Service	Hawaii	2001	17	Improve quality of care	-	X	X	X	X			Unknown	Improvement in process indicators, minor increase in patient satisfaction	Unknown
	Hospital Agreement Incentive Program	Michigan	2001	86	Improve quality of care	-	X	X	X	X			4%	Improvement in process indicators	Unknown
	Hospital Quality Incentive Demonstration Project	Nationwide	2003-2009 (ceased)	262	Improve quality of care	5		X	X			X	2%	Mixed improvement in process indicators, no improvement in outcomes, no sustained effect	Increase in health disparities, unreasonable punishment of hospitals serving disadvantaged population groups
	Non-Payment for Hospital-Acquired Conditions	Nationwide	2008	unknown	Reduce prevalence of patients with hospital-acquired conditions	13 in 2009 (changes annually)		X			X		0.001%	Improvement for some HACs, no improvement for others	Increase in public attention for HACs
	Hospital Value-based purchasing (HVBP) incentive payment	Nationwide	2011	4 783	Improve quality of care in selected interventions	3 (changes in 2016 and 2017)		X	X		X	X	Up to -1-2%	No effect detected yet	Unreasonable punishment of hospitals serving disadvantaged population groups
	Hospital readmission reduction program (HRRP)	Nationwide	2012	unknown	Reduce readmission rates	5			X			X	Up to -3%	Early research suggests decrease in readmissions	Increase in public attention for readmissions, may increase disparities
	Hospital-Acquired Condition Reduction Program (HACRP)	Nationwide	2015	unknown	Reduce prevalence of hospital-acquired conditions	10 in 2015 (changes in 2016 and 2017)		X				X	Up to 1%	Not evaluated yet	Not evaluated yet, may increase disparities

Australia:[21-24]; Canada:[25-28] ; Denmark: [29] ; France:[30-32] ; Israel: [33, 34]; Italy: [35, 36]; Japan: [37-39]; Luxemburg: [40-43]; Norway: [44]; South Korea: [45-48]; Sweden:[49-52] ; Turkey: [53-57], England: [58-66], United States: [67-81]

3. Results

3.1. *Aims*

We identified 30 P4P programs in 14 countries. They covered a wide range of aims. At one end of the spectrum, we found P4P programs that focused on specific targets. Four of the programs focused on meeting defined targets for waiting times. The Italian P4P program, for example, was intended to reduce waiting times for hip replacements among patients aged 65 and above [82]. This aim is based on the assumption that shorter waiting times lead to improved outcomes. Several other programs aimed at improving the quality of care in patients with selected conditions. The Japanese P4P program, for example, targeted stroke and was designed to improve the delivery of care and functional outcomes in patients with that condition [37]. In contrast, the South Korean *Value-incentive payment (VIP)* targeted acute myocardial infarction (AMI) and Caesarian sections, aiming to improve the overall performance of hospitals while reducing variation among them for these two conditions. Policy makers in South Korea chose these two conditions for this P4P program because the country had ranked among the worst for their care in an OECD comparison [83].

At the other end of the spectrum, we identified P4P programs intended to improve the quality of care for a range of conditions and for the inpatient sector as a whole. The English initiative called *Advancing Quality*, for example, aimed to reduce costs and mortality for 14 conditions. These were chosen because the NHS Northwest region had performed poorly on them in a regional comparison, a result that could largely be explained by a high degree of deprivation in that geographic area. Furthermore, those who designed the program had identified these as conditions that would particularly benefit from incentive payments [84]. The high volume of cases and the ability to measure quality of care through indicators also played a role in their selection [64]. Programs that focused more strongly on the overall performance of hospitals were the Swedish *målrelaterat ersättning* [Quality-based remuneration] and the *Incitants qualité* [Quality stimulators] in Luxembourg. The latter was part of a comprehensive reform of

the Luxembourg hospital system in 1998, which aimed at budget containment, reforming the organization and infrastructure of inpatient care nationally, increasing transparency and creating a legal framework for quality of care. This was also true of the *HVBP* in the US, which formed part of the Affordable Care Act [85].

3.2. Selection of indicators

The indicators used in the P4P programs we identified were either of a more general nature or tied to a specific condition (see table 2 below). They can be grouped into structural, process and outcome indicators. Structural indicators were rarely used. In cases where they were, they referred to specific characteristics of a hospital or of patients. Examples of the former are staffing ratios and efficiency (defined as average spend per patient); an example of the latter is greater disease severity, which was used as an indicator in the French and Japanese P4P programs.

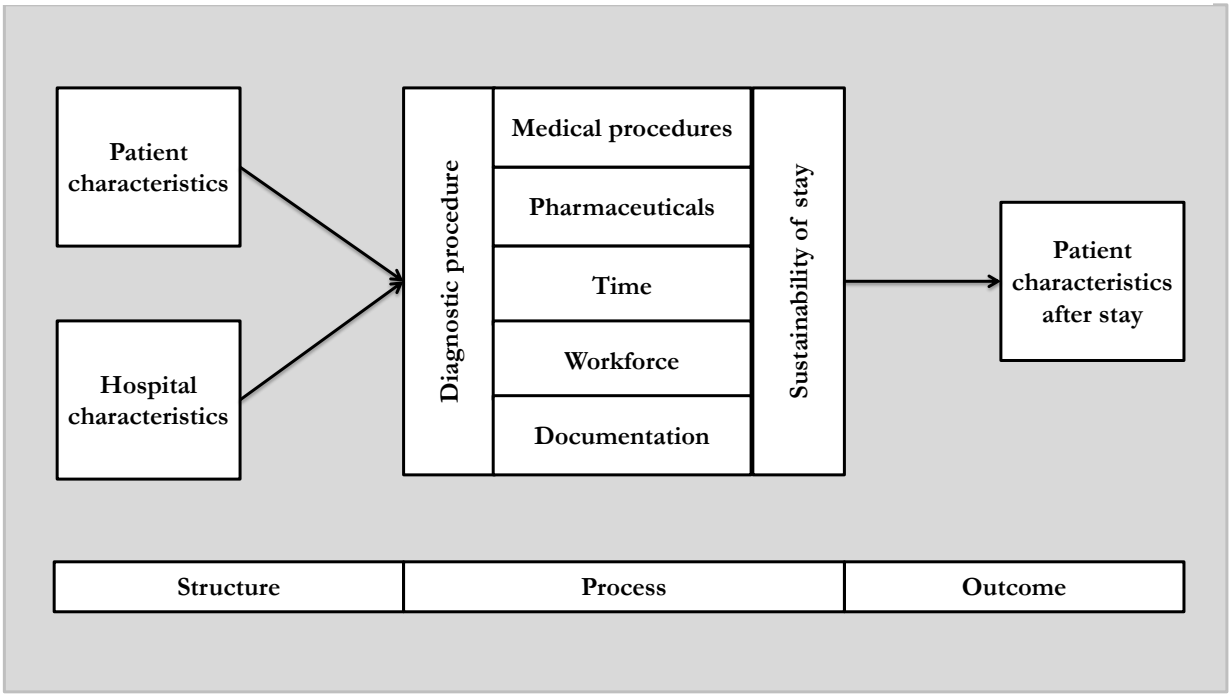
In contrast to structural indicators, process indicators were used in all of the P4P programs. These indicators targeted different stages in the delivery of care. Some of the process indicators were chosen to improve diagnostic procedures and came with a time target attached. The English *Advancing Quality* and *Best Practice Tariffs* programs, for example, required a CT scan or an MRI within 24 hours of admission for stroke patients. Another set of indicators focused on the delivery of care during the inpatient stay. Process indicators during inpatient treatment can be divided into five categories as shown in figure 1, namely those that incentivize hospitals to:

- follow predefined treatment pathways (for example, blood and urine tests at regular intervals for diabetes in the *Advancing Quality* initiative in England);
- deliver appropriate and timely pharmaceutical treatment (for example, aspirin intake within 30 minutes after arrival for acute myocardial infarction, as in the *Premier HQID* in the United States);
- meet specific time targets (for example, hip replacement surgery within 48 hours of admission, as in the P4P program in Lazio, Italy) [25]

- meet certain hospital workforce requirements (for example, review by a senior clinician or critical care team within 4 hours of arriving at the hospital, as in the *Advancing Quality* initiative in England)
- improve documentation (for example, submitting patient information to national registries as is the case in Sweden)

Some P4P programs went beyond the inpatient stay and rewarded hospitals for processes intended to improve post-discharge care and avoid readmission. For example, in the *Best Practice Tariffs* program in England, the cause of low blood sugar must be discussed with the patient before discharge. Furthermore, the patient is discharged with a written care plan, which is sent to the GP, and is offered a structured education program within three months after discharge.

Figure 1: Intervention points of indicators



Source: Own representation by the authors.

Outcome indicators were used less frequently than process indicators, but more commonly than structural indicators. They were most prevalent in P4P programs in the US. The two dominant indicators to assess health outcomes were the 30-day mortality

rate and the rate of hospital readmissions. In very few cases, outcome indicators went beyond the 30-day period, such as the Swedish P4P program, which measured reoperation rates two years after hip replacement surgery.

Table 1: Overview of condition-specific indicators

Condition	Type	Content[Country]
Acute conditions		
AMI	P	Aspirin prescribed at arrival$_{ROK, ENG1, USA3, USA4}$/ at discharge$_{ROK, ENG1, USA2, USA3}$ ACEI/ARB (for LVSD) (at discharge)$_{ENG1, SWE, USA2, USA3}$ Beta blocker at arrival$_{USA3}$/ at discharge$_{ROK, USA2, USA3}$ Adequate prescription of medication$_{FRA}$ Coronary circulation X-ray$_{SWE}$ ST elevation after reperfusion$_{SWE}$ Patients with P2Y12-receptor antagonist$_{SWE}$ Primary PCI received within 90 minutes/ 120 minutes of hospital arrival$_{ROK, USA3, USA4}$ Smoking cessation advice/ counseling$_{USA2, USA3}$ Lipid-lowering therapy (on discharge)$_{SWE, USA2}$ Fibrinolytic received within 30 minutes/ 60 minutes of hospital arrival$_{ROK, USA 3, USA 4}$ Diet counceling after AMI$_{FRA}$
	O	Inpatient mortality rate (JCAHO risk adjustment)$_{USA3}$/30-day mortality rate$_{ROK, USA3}$ Rate of readmission within 30 days$_{USA3}$ AHRQ patient safety indicators (combined to a complication index)$_{USA3}$
CABG	P	Aspirin prescribed at discharge$_{USA3}$ CABG using internal mammary artery$_{USA3}$ Prophylactic antibiotic received within 1h prior to surgical incision and selection for surgical patients within 48h of surgery end time$_{USA3}$
	O	Inpatient mortality rate (JCAHO risk adjustment)$_{USA3}$/30-day mortality rate$_{USA 3}$ Postoperative hemorrhage or hematoma/ psychologic and metabolic derangement$_{USA3}$ AHRQ patient safety indicators (combined to a complication index)$_{USA3}$
COPD(acute)	P	Oxygen levels in blood checked and targeted prescribed within 48h of hospital arrival$_{ENG1}$ Steroids appropriately administered within 4h of hospital arrival$_{ENG1}$ Inhalers administered within 4h of hospital arrival$_{ENG1}$ Smoking cessation councelingg$_{ENG1}$ Offer of referral to pulmonary rehabilitation$_{ENG1}$ Review of inhaler techniques$_{ENG1}$ Written self-management plan$_{ENG1}$ Home oxygen therapy assessment$_{ENG1}$ Spirometry referral made$_{ENG1}$ Plans for managing end stage of the disease$_{ENG1}$
Diabetes (actue)	P	Referred to diabetes specialist team at admission/ seen within 24h prior to discharge$_{ENG2}$ Blood glucose level checked within 30 minutes of hospital arrival$_{ENG1}$ Foot inspection within 24h of hospital arrival$_{ENG1}$ Specific assessment carried out at recommended intervals$_{ENG1}$ Blood and urine tests at regular intervals$_{ENG1}$ Fluids and insulin via IV drip within 60 minutes of DKA detection$_{ENG1}$ Reviewed by senior clinicians within 12h of DKA detection$_{ENG1}$ Foot ulcer description within 4h of detection$_{ENG1}$ Antibiotics within 6h of foot ulcer detection$_{ENG1}$ Referred to hospital foot care team within 24h and seen within 72h of referral$_{ENG1}$ Quick acting carbohydrates given within 15 minutes of low blood sugar detection$_{ENG1}$ Blood glucose monitored after administration of quick acting carbohydrates$_{ENG1}$ Escalation of care if blood glucose remains low after 45 minutes of quick acting carbohydrates being administered$_{ENG1}$ Cause of episode of low blood sugar discussed with patient before discharge$_{ENG1, ENG2}$ Discharged with written care plan which is copied to the GP$_{ENG2}$ Offer of structured education within 3 months after discharge$_{ENG2}$
Heart Failure	P	Evaluation of LVS function$_{ENG1, USA3, USA 4}$ ACEI/ ARB for LVSD$_{ENG1, USA3, USA4}$ Detailed discharge instructions$_{ENG1, USA3}$ Smoking cessation advice/ counseling$_{ENG1, USA3}$
	O	Inpatient mortality rate (AHRQ IQI)$_{USA3}$/ 30-day mortality rate$_{USA3}$ Rate of readmission within 30 days$_{USA3}$AHRQ patient safety indicators (combined to a complication index)$_{USA3}$
Pneumonia	P	Appropriate initial antibiotic selection$_{ENG1, USA 3}$/ antibiotics within 6h after arrival$_{ENG1}$ Blood culture performed in emergency department prior to first antibiotic received in hospital$_{ENG1*, USA3, USA4}$ Influenza vaccination$_{USA 3, USA 4}$/ pneumococcal vaccination$_{USA 3, USA 4}$ Oxygenation assessment$_{ENG1, USA3}$/ assessment of severity of pneumonia ("Curb-65")$_{ENG1}$ Smoking cessation advice/ counseling$_{ENG1, USA3}$

	O	Inpatient mortality rate (AHRQ IQI)$_{USA\ 3}$/ 30-day mortality rate$_{USA3}$ Readmission within 30-days rate$_{USA3}$ AHRQ patient safety indicators (combined to a complication index)$_{USA3}$
PPH	P	Minimum clinical supervision in delivery room$_{FRA}$ Prevention of postpartum hemorrhage$_{FRA}$
Sepsis	P	Screening for sepsis within 2h of hospital arrival$_{ENG1}$ Blood tests/ test for level of lactic acid within 3h of hospital arrival$_{ENG1}$ Antibiotics within 3h of hospital arrival$_{ENG1}$ Second liter of IV fluids/ oxygen therapy within 4h after hospital arrival$_{ENG1}$ Fluid balance chart within 4h of hospital arrival$_{ENG1}$ Review by senior clinician or critical care team within 4h of hospital arrival$_{ENG1}$
Stroke	S	>20% are severe cases$_{JAP}$
	P	Direct admission to, and majority of length of stay, in stroke unit$_{ENG2}$ Antiplatelet/ anticoagulant therapy (within 48h)$_{AUS\ 1,\ SWE}$ Thrompolysis assessment/ therapy$_{ENG2,\ SWE}$ Dysphagia screen within 24h$_{AUS\ 1}$ Time of recording of symptoms$_{FRA}$ Treatment by specialist nursing, specialist neuro-intensivist care, qualified clinicians$_{ENG2}$ Admission to stroke unit (within 48h hours after arrival)$_{ENG1,\ ENG2,\ SWE}$ Test of ability to swallow$_{ENG1}$ CT or MRI within 24h$_{ENG1,\ ENG2}$ Blood thinning medication within 24h$_{ENG1}$ Weighting during stay$_{ENG1}$ Assessment of movement/ of ability to carry out day-to-day tasks within 72h$_{ENG1}$ Share of patients registered in national stroke registry$_{SWE}$
	O	>30% show improvement in activities of daily living/ functional recovery at discharge$_{JAP}$ > 60% of all stroke patients discharged into the community$_{JAP}$
Major trauma	P	If Injury Severity Scale>8: Patient treated in a major trauma center (MTC)$_{ENG2}$ Transferred from trauma unit to MTC within 2 days if transferred as a non-emergency case$_{ENG2}$ Tranexamic acid within 3h of injury if applicable$_{ENG2}$ Trauma Audit and Research Network data completed within 25 days of discharge$_{ENG2}$ If Injury Severity Scale >16: Received by trauma-led team with consultant within 5min$_{ENG2}$ Head CT within 60min if no emergency surgery/ interventional radiology applicable$_{ENG2}$ Transferred from trauma unit to MTC within 2 days if transferred as a non-emercency case$_{ENG2}$
Elective/ chronic conditions		
Cataract	P	Initial diagnosis of cataract in primary care$_{ENG2}$ Confirmation of diagnosis, listing for surgery and pre-operative assessment$_{ENG2}$ Cataract removal on a day case basis$_{ENG2}$ Follow-up of surgery 2 weeks after surgery by nurse, optometrist or ophthalmologist$_{ENG2}$ Review at 4 to 6 weeks by optometrist$_{ENG2}$
COPD	P	Pulmonary rehabilitation program according to standards (min. 8 weeks for exercise training, multidisciplinary education, at least 2 exercise session per week, evaluation of quality pre- and post-program, exercise capacity pre-and post-program$_{AUS1}$
C-section	P	Ratio between observed and expected Caesarian section rate (based on 15 clinical risk factors)$_{ROK}$
Dementia	P	Nutritional needs assessment within 5 days of admission$_{ENG1}$ Physical health/ initial pain assessment within 7 days of admission$_{ENG1}$ Cognitive ability/ depression assessment within 14 days of admission$_{ENG1}$ Functional capacity assessment before leaving hospital$_{ENG1}$ Patient focused care plan before leaving hospital$_{ENG1}$
Diabetes	O	Share of Type 1 patients having reached treatment goal of HbA1c <5.2/7.3$_{SWE}$ Share of Type 1 patients having reached treatment goal of blood pressure <139/80mm Hg$_{SWE}$ Share of Type 1 patients having reached treatment goal of LDL-cholesterol <2.5 mm/L$_{SWE}$
Dialysis	P	85% of patients receive dialysis through functioning arteriovenous fistula$_{ENG2}$
	O	Share of patients with arteriovenous fistula or arteriovenous graft$_{SWE}$ Share of patients who have achieved blood pressure treatment goals$_{SWE}$ Share of patients having reached treatment goal of Kt/V>2 of dialysis dose$_{SWE}$ Share of patients receiving dialysis at home$_{SWE}$
FHF	P	Surgery within 36h after admission$_{ENG2}$ Joint admission by consultant geriatrician and consultant orthopedic surgeon$_{ENG2}$ Assessment protocol agreed by geriatric medicine, orthopedic surgery and anesthesia$_{ENG2}$ Postoperative geriatrician-directed multi-professional rehabilitation team$_{ENG2}$ Fracture prevention assessment and Abbreviated Mental Tests pre/ post-surgery$_{ENG2}$

Hips & knees	P	Surgery within 24h$_{SWE}$/ within 48h for patients aged 65 and above$_{ITA}$ Prophylactic antibiotic within 1h prior to surgical incision/ discontinued within 24h of surgery end time$_{USA\ 3}$ Appropriate venous thromboembolism prophylaxis within 24h$_{USA\ 3}$ Pre- and postoperative period Post-operative hemorrhage or hematoma/ physiologic and metabolic derangement$_{USA\ 3}$ Coverage of patients in national quality registry for hip replacements$_{SWE}$
	O	Share of patients with surgery 2 years after hip TEP$_{SWE}$ Readmission within 30 days to acute care inpatient rate$_{USA3}$ AHRQ patient safety indicators (combined to complication index)$_{USA3}$
Psychosis	P	Risk assessment within 30 days of acceptance$_{ENG1}$ Assignment of care coordinator$_{ENG1}$ Antipsychotic medication review$_{ENG1}$ Calculation of duration of untreated psychosis and measurement of symptom severity (using PANSS)$_{ENG1}$
Schizophrenia	P	Patients seen by a community mental health professional within 7 days after discharge from same district mental health service provider$_{AUS1}$ Recording antipsychotic injection (depot) medication on iPharmacy$_{AUS1}$

Note: Countries appear in alphabetical order. Within countries, programs appear in historic order. AUS 1: Queensland CPIP; ENG1: Advancing Quality; ENG2: Best Practice Tariff; ITA 1: Waiting-time strategy in Lazio; JPN: Japanese stroke P4P; ROK: Value-incentive payment; SWE: Västa Gotaland; USA 1: BCBS of Michigan, US 2: Hawaii, US 3: Pemier HQID, US 4: Value-best practicing program.
Some programs include additional conditions that can also be treated in the outpatient sector, such as non-acute diabetes care. We have not included these conditions in such cases.

3.3. Financial rewards

When designing financial rewards, policy makers have several points at which they can intervene:

- What to reward: Broadly speaking, policy makers can choose between rewarding providers based on some aspect of the care they deliver for a specific condition or a group of conditions.
- Who to reward: Here, again, two options are possible. The first is to reward hospitals based on an absolute threshold. In such cases, every hospital that meets a certain quality threshold is rewarded and/or penalized if it falls below it. The second option is to reward hospitals in relative terms. This can be done based on status quo performance, or changes over time (in comparison to other hospitals or to the hospital's own performance in an earlier period).
- Whether to use carrots or sticks: Incentives can take the form of bonuses, penalties, withholds or a combination thereof. They can also be lump-sum payments, or a percentage of a hospital's budget or of the payment a hospital receives for treating a specific condition.

- How much to reward: In the P4P programs identified in our research, the financial incentives often amounted to approximately 0.1% of a hospital's budget or less, and never exceeded 4%.

3.3.1. What to reward?

The decision about what to reward depends on the aim of a P4P program. Broadly speaking, there were two options available in the OECD countries with P4P program. The first involved granting rewards based on some aspect of the care delivered for a specific condition or by a clinical department. The P4P program in Lazio, Italy, was an example of the former, and the Canadian P4P programs were an example of the latter. In both cases, waiting times were used as an indicator. Taking a somewhat different approach, the Japanese P4P program targeted the care delivered for a specific condition, stroke, but measured this using three indicators instead of one. In its most comprehensive form, the *Best Practice Tariffs* in England covered 64 conditions, but these were not combined to a joint score. Instead, a hospital could decide for each condition and for each patient as to whether to treat him or her under standard conditions or the best practice tariff.

The second option involved granting rewards for multiple aims simultaneously. Programs that aimed to improve the quality of care for several conditions together, or to improve the overall quality of care in a hospital, tended to rely on an aggregate score. In contrast to the first option described above, this type of program followed an all-or-nothing approach because hospitals could not decide to treat some patients under the P4P quality criteria other patients under standard conditions. Instead, hospitals had to meet minimum requirements for all dimensions. In programs such as *incitants qualité* in Luxemburg, *Advancing Quality* in England, the *HVBP* in the US and the *målrelaterat ersättning* in Sweden, various indicators added up to an overall score, which was used to calculate rewards. In some cases, weights were applied. The *HVBP*, for example, gave preference to outcome indicators. Indicators covered the domains of

"outcome", "efficiency", "patient experience" and "process" and were weighted with 40%, 25%, 25% and 10%, respectively.

3.3.2. Who to reward?

When categorizing hospitals as high or low performers for the purpose of allocating rewards, two methods were most prevalent. The first of these involved rewarding hospitals based on their *absolute* performance. In such cases, hospitals either received a bonus if they exceeded a certain performance threshold or incurred a penalty if they fell below it. In the P4P program in British Columbia, for example, hospitals received a bonus for each patient who left the emergency department within a pre-specified amount of time [25]. Similarly, in the P4P program in Lazio, Italy, hospitals received the full DRG payment for patients aged 65 or above who underwent hip replacement within 48 hours after hospital admission [82]. It is important to note that absolute scores can be simple or stepped. In the latter case, targets are tightened at regular intervals, making it increasingly difficult for hospitals to meet them. The idea is to encourage continuous improvement. In the case of simple absolute scores, the targets do not change over time. The second method involved rewarding hospitals based on their *relative* performance. This took three forms:

- The "tournament" [62, 65] or "top-/worst-performer award": In this setting, the current performance of a hospital was compared to that of other hospitals. Only the best hospitals received a bonus and/or only the worst ones incurred a penalty. This was applied in the first phase of the *Premier HQID* program in the US and *Advancing Quality* in England, and was still being used in the *VIP* in South Korea.
- A reward based on the performance of a hospital compared to that of other hospitals over time, something referred to as an "achievement" award [62, 65, 86]: This was generally expressed as the performance of a hospital compared to the median performance of all hospitals in a prior period. In *Advancing Quality*,

the comparison period was 12 months, whereas it was 24 months in the *Premier HQID*.

- An "improvement" award [87]: This can be understood as a reward based on a hospital's current performance compared to its own performance over time, such as in the *HVBP*.

P4P programs that rewarded hospitals on a relative basis blended these three methods in various ways. The *Premier HQID* and *Advancing Quality* programs combined top-performer awards with an attainment award. In the latter program, receiving a top-performer award was conditional on having received an attainment award, whereas this was not the case in the *Premier HQID* program. In both programs, awards could be added on to one another. In the *HVBP* in the US, however, hospitals could receive either the attainment or the improvement award, but not both. Additionally, in the *HVBP*, there was no top-performer award. To further maximize the incentive structure, the *Premier HQID* also granted additional awards if hospitals were eligible for both awards. In *Advancing Quality*, those that performed particularly well in the attainment dimension (meaning that they were among the top 25% of those having improved their performance compared to the year before) could receive an additional bonus.

Figure 2: Award structure of three P4P programs

	Premier HQID	Advancing Quality	VBP incentive payment
Top-performer „Am I the best?"	X*	X	
Attainment „Is my improvement better than the one of my peers?"	X	X + X	X **OR**
Improvement „Am I better than I was before?"			X

Source: Own representation by the authors.

*= conditional on being eligible for the attainment award.

3.3.3. *Sticks or carrots?*

P4P programs can also differ in terms of whether they use payment withholds, bonuses, penalties, or a combination of these to incentivize certain behaviors. Several factors can shape this decision, including financial constraints, the desired magnitude of a program's effect, acceptance of the program by hospitals, and the cultural, or normative, understanding of a P4P program.

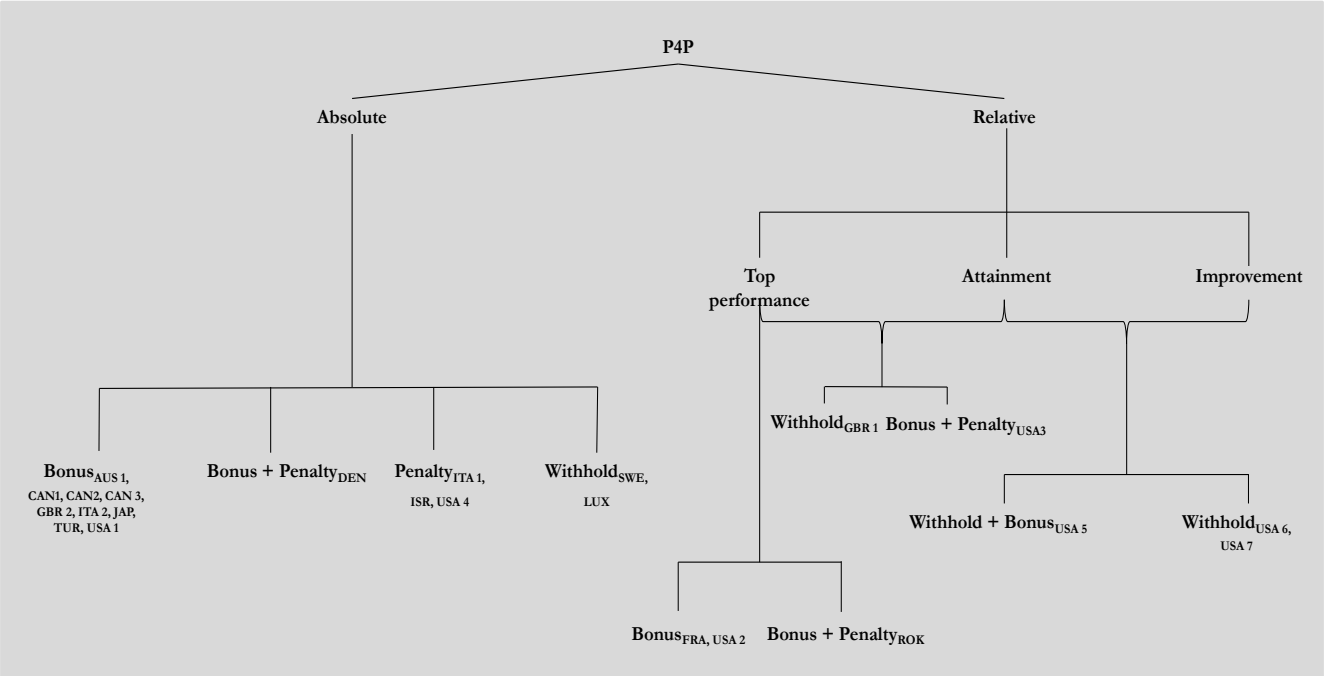
The first factor is financial constraints. In particular, policy makers must decide between programs that are budget neutral and those that incur additional costs. If a P4P program needs to be budget neutral due to fiscal constraints, policy makers have a range of design options.

- Payment withholds: This option involves withholding payments prospectively and releasing them retrospectively only if a hospital meets certain quality criteria. This was the case in the P4P programs in Sweden and Luxemburg, and with the English *Advancing Quality* program.
- Payment withholds combined with a redistribution mechanism: The *HVBP* in the US was an example of this. Here, all hospitals were subject to a withhold payment. The financial resources resulting from this were subsequently redistributed to all hospitals that scored sufficiently high on the attainment or improvement score. The amount of the incentive was adjusted based on the financial resources to ensure budget neutrality. A blend of bonuses and penalties: Alternatively, policy makers might consider blending bonuses with penalties as was done in South Korea's *VIP*. In this case, the worst performers received a deduction of up to 2% whereas the best performers received a bonus of 2% of the payments by the National Health Insurance [83].
- Penalties alone: If budgets are particularly tight, policy makers may decide to use penalties alone. In such cases, hospitals' DRG payment or total budget is reduced if they do not satisfy certain quality requirements. An example of this can be found in the waiting times program in Lazio, Italy. In essence, this type of penalty does not differ greatly from a withhold payment, as both approaches

involve either reducing a hospital's DRG payment or total budget, or leaving these unchanged. One difference, however, is that withholds are issued prospectively, whereas penalties are imposed retrospectively.

If budget neutrality is less of a concern, policy makers may choose to use bonus payments alone. This approach was frequent among the programs we identified in our research. Examples include the French *IFAQ* program and the Canadian P4P programs for emergency departments.

Figure 3: Design elements of financial incentives in P4P programs



3.3.4. How much to reward?

Across the P4P programs, the proportion of revenue a hospital could generate through P4P was consistently low. Very often this was approximately 0.1% of a hospital's budget or less, and it never exceeded 4%. The financial burden of such programs for governments or health insurers was therefore low compared to their total annual hospital expenditure.

Of course, when looking at the incentives on a per condition basis, the amount can appear quite substantial. In the Italian P4P program, for instance, the DRG payment for a hip replacement can be reduced by up to 50% if waiting times exceed 6 days [78]. Similarly, in the British Best Practice Tariffs program, the performance-based component can comprise up to 50% of the payment a hospital can receive for the respective condition – a share which had risen to this level through continual increases since the program's inception. However, the sums in question remained negligible compared to the total budget. In 2011/12, English hospitals received GBP 71 million through Best Practice Tariffs. This compares to expenditure of GBP 532 million for the relevant procedures, and to GBP 30 billion of total expenditure for the entire acute care sector in England [57]. In England's second P4P program, Advancing Quality, the financial volume amounted to GBP 7.1 million, of which GBP 5 million were dispersed as bonuses. In the second year, the volume increased to GBP 10.6 million, representing 0.1% of the total budget of the NHS North West Strategic Health Authority. After changes in the reward structure from bonus payments to payment withholds, the maximum withhold could amount to GBP 5 million [58, 60].

Payments were similarly low in other countries in relation to total hospital expenditure. In Japan, hospitals that met the quality requirements received a bonus payment of Yen 17 200 (approx. EUR 130). In 2010, 85% of participating hospitals received bonus payments [33]. In the P4P program in South Korea, payments were similarly low. In 2009, 21 out of 43 hospitals received rewards totaling 453 million Korean WON (approx. EUR 373 200), and in 2010, 26 out of 43 hospitals received rewards totaling 404 million WON (approx. EUR 332 500) [85]. In British Columbia, the P4P program granted CAD 21.3 million for the treatment of 11 048 patients to the 14 participating hospitals for the fiscal year 2011/12 [86]. This amounted to 2.7% of the budget for patient-focused funding and 0.3% of the total budget devoted to acute care. For the fiscal year 2012/13, this amount rose only slightly to CAD 25 million, representing a share of 2.5% of the budget for patient-focused funding. Payments granted by Ontario were of a similar volume. In 2013/14, the financial volume of this

program amounted to CAD 93 million for 74 participating hospitals [87]. Programs in the US were also of low cost. In the first year, the Premier HQID program's financial volume amounted to USD 8.85 million and was distributed to 123 of the 248 participating hospitals. This amount decreased to USD 8.5 million for 115 hospitals in the second year and 7 million for 112 hospitals in the third year. The average bonus awarded to each hospital was USD 71 960 per year, and ranged USD 914 to USD 847 227 [88].

3.4. Impact of the various P4P programs

Of the 30 P4P programs identified in our research, very few had been evaluated. When evaluations had been undertaken, they were characterized by limitations, including small sample sizes, the lack of a control group, or the presence of confounding factors, such as the introduction of public reporting alongside the implementation of the P4P program [91]. The results of these evaluations suggest that the effects of the P4P programs ranged from none to moderately positive, did not meet government expectations and were not sustainable [88].

3.4.1. Program effects

Among the more narrowly focused and straightforward initiatives, the Canadian P4P programs showed improvements that were modest but below expectations. Moreover, the results of the evaluations were heterogeneous, and the causalities were unclear. In British Columbia, hospitals in the region of Vancouver Island were able to maintain or even increase the share of patients who met waiting time targets despite a rise in the number of emergency department visits. This was not true, however, for hospitals in the region of Fraser, where waiting times increased during the study period.

The results were also mixed for hospitals in Ontario. Evaluations of these hospitals found that, overall, the 90th percentile and median length of stay in the emergency department, as well as time to physician assessment, improved significantly [92]. Furthermore, there were no unintended consequences in terms of short-term admissions, mortality or readmissions [92]. When the participating hospitals were

compared to the control group of non-participating hospitals, however, the improvements were much smaller or non-existent. Moreover, in a few cases, hospitals in the control group outperformed those in which a P4P program had been implemented [92]. For both British Columbia and Ontario, there were various confounding factors. In Ontario, a public reporting program and changes in the remuneration system were introduced several months before the P4P program [92]. Unfortunately, it is not possible to understand why some hospitals improved their performance and others did not, as it is not known whether or how hospitals restructured their emergency departments [90].

An evaluation of the Japanese P4P program for stroke patients also yielded mixed results. While process indicators appear to have improved in some hospitals, for example through an increase in the intensity of treatment, no improvements were seen in the functional recovery of stroke patients after their hospital stay [34]. However, there was no improvement in the functional recovery of stroke patients after hospital stay even for patients who received more intensive treatment. Unfortunately, it was not clear if the latest modifications to the program, which took place in 2012, had had an impact on hospital performance. In the US, the results of evaluations of P4P programs were similarly mixed. The programs that had been evaluated at the time of our research were the Blue Cross Blue Shield of Michigan Participating Hospital Agreement Incentive Program and the Hawaii Medical Service Association Hospital Quality Service and Recognition P4P Program. While the evaluations of these programs suggested that improvements had been made, it was unclear to what extent these could be attributed to financial incentives or other factors [63, 65, 93, 94]. The effects of the Premier HQID program, which had received wide attention and became the blueprint for several later P4P programs, have also been contested [91, 95, 96].

Overall, the evaluations undertaken of US programs to date have generally shown an improvement in process indicators but not in patient outcomes. Lindenauer et al. found that process indicators improved in participating hospitals compared to a control group

[88]. Further studies showed a slightly positive association between process indicators and health outcomes for acute myocardial infarction, congestive heart failure and pneumonia [97, 98]. In cases for which there was evidence of early positive effects, these did not appear to be sustainable, however. Lindenauer et al. found that after five years of the HQID program, there were no differences between participating hospitals and their non-participating counterparts [88]. From 2003 to 2006, the HQID used top-performer awards only. According to the analysis by Ryan et al., hospitals with more deprived patients were systematically disadvantaged in the period from 2003 to 2006. These disparities declined following a system change that took place in 2006, thus improving the fairness of the program [83]. From that point onwards, the program also granted attainment awards and a combination of bonuses and penalties. Based on this, Ryan et al. posit that the HVBP has a promising outlook as it is modeled in a similar way [14].

Similar criticisms were made of the Advancing Quality initiative in England, which was modeled on the Premier HQID program. Evaluations of the first two waves of Advancing Quality concluded that it was a success. Over the first 18 months, absolute mortality decreased by 1.3 percentage points among patients with pneumonia, heart failure and acute myocardial infarction. This translated into a relative reduction of 6% and 890 lives saved. The financial volume of 13 million pounds generated 5 200 QUALYs and savings of 4.4 million pounds due to reductions in length of stay [99]. The cost-effectiveness of this program is much below the threshold of NICE. It is assumed that it has saved 17 million pounds within the first three years. The initially promising findings of the Advancing Quality program do not appear to be sustainable, however. A first look at the program's long-term effects indicates that the initially positive effects appear to disminish over time [100]. Forty-two months after the program was introduced, the reduction in mortality in non-participating hospitals was greater than in the participating ones. Conversely, the decline in mortality for conditions which were not part of the P4P program was greater in P4P hospitals than non-P4P programs. Due to the frequent changes in the incentive structure of the program, it is not clear

whether the initially promising results would have been maintained if the design had remained unchanged. The continuing development of the South Korean VIP, which also mirrors the Premier HQID, may provide some insight in this regard. At the time our review was completed in December2015, it had succeeded in improving the overall delivery of care while reducing variation [85, 101]. However, these results must be interpreted with caution as the program had not been evaluated in a thorough way and had never been compared to a target group.

Evaluations of one of the most recent and complex programs, the HVBP, yielded mixed results. Ryan et al. failed to find a correlation between the incentive payment and improvements in clinical process or patient experience compared to a control group in the initial implementation period [72]. Additionally, the effect of the HVBP did not vary depending on the hospital's initial performance in these two domains. The authors did, however, find improvements before the P4P program had begun. Hospitals that expected to be subject to the HVBP began to improve their clinical performance about three years beforehand [72]. Whether this was driven by the expectation of the incentive payment remains unclear. This pattern was not observed for patient satisfaction [95]. These results were supported by Spaulding, Zhao and Haley, who were unable to identify a correlation between the total performance scores and patient safety and quality in the domain of hospital-acquired conditions [102]. It has to be taken into account that the results originate from the initial phase of the HVBP program. The composition of the aggregate score based on which incentives are granted has changed, and the financial amount that is withheld and redistributed has increased since that. It is therefore not yet possible to draw final conclusions from these evaluations.

3.4.2. Side effects and unintended consequences

Information on the side effects of P4P programs is even scarcer than evaluations of their direct effects. When available, such information points to several potential side-effects that need to be taken into account when implementing a P4P program.

To begin with, P4P programs may incentivize hospitals to actively change their delivery structure. If a hospital performs well in the treatment of a condition that is subject to a P4P program, but poorly in the treatment of a condition for which it receives no bonus or might even receive a penalty, it can be expected to shift the delivery of care towards the treatment at which it excels. Furthermore, a P4P program might increase adverse selection and thus lead to unacceptable variation in access to care. For example, it has been argued that the HVBP disproportionally penalized hospitals with disadvantaged patient populations [103, 104]. This sets a clear incentive to give preference to low-risk patients over high-risk ones, such as older or deprived patients, or patients with various co-morbidities whose treatment course is less easy to predict. Also the Japanese P4P program had come under criticism for these reasons, but the evidence for adverse selection in this case is unclear [105]. After the P4P program was introduced, the share of patients with better functional scores at admission increased which may point at adverse selection. At the same time, the government introduced a new type of nursing home for severe stroke patients. It may therefore also be that severe stroke patients were being treated by different providers [105]. Similarly, the results of a survey of health care professionals in South Korea suggest that the program there was prone to unintended consequences, with adverse selection effects and quality skimping in areas not subject to the program being of particular concern [43].

Besides such active changes in their service delivery structure, hospitals can also be affected in a passive way by factors that are out of their control. Several P4P programs resulted, for example, in administrative difficulties. While these might not impact on the delivery of care per se, they can unreasonably penalize hospitals. In Ontario, for

example, incentives were granted with massive delays. Furthermore, the allocation was questionable. In the first year, only three of 23 hospitals had met the government targets. Several hospitals that had failed to meet targets in the first year received even greater incentives the following year. The hospital that had performed worst in the first year received the greatest amount of incentives in the second year [106]. Similar issues occurred with the BPTs in England. There, hospitals received either a greater amount than that to which they were entitled to, or one that was less than that which they should have received. Providers and commissioners reported limited knowledge about the relationship between the quality provided and potential financial rewards, poor data quality and financial constraints as problems they encountered in the implementation of BPTs. In some cases, commissioners did not pay BPTs because their budget did not allow them to do so, even in cases where the hospital might have been eligible [89]. In line with that, if not designed carefully, a P4P program can punish hospitals for factors that are not in their control. If the data used to determine rewards is not sufficiently adjusted for patient characteristics, this might unreasonably punish hospitals for serving a disadvantaged population. For the vast majority of P4P programs, side effects have not been investigated. In other cases, they are mentioned as a potential concern or assumed to take place, but not investigated [33]. If evaluations on side-effects are undertaken, these studies generally confirm their occurrence. The HVBP, for example, faced criticism that it unjustifiably punished large and safety-net hospitals [103].

4. Discussion and policy recommendations

This review indicates that P4P has become an integral part of the remuneration of hospitals in the OECD with a high degree of heterogeneity in terms of their design. Their evaluations, though, show that the programs yield modest, short-term improvements at best. Thus, following a first wave of enthusiasm, it has been put into question whether P4P is the most effective way to use public money [58, 107]. This critique is not singular to the health care market. In other policy fields, it is equally

contested, its effects are heavily disputed, findings report mixed results and unintended consequences might arise if the programs are not designed in a careful way [108-110].

Several critiques prevail. To begin with, some researchers and policy makers questioned whether financial incentives would have any effect, at all. A long-term evaluation of Advancing Quality showed that the control group eventually showed greater improvements in quality than the intervention group of hospitals which participated in the P4P program [58]. In line with that, Jha et al. and Ryan et al. could not find any long-term effect in hospitals participating in the Premier HQID program compared to their non-participating counterparts [107, 111] and the same may eventually hold true for the HVBP [72, 111]. Additionally, P4P has also come under fire for potentially crowding out motivation [112]. Selected papers have pointed into the direction that P4P may be suspected to create extrinsic motivation rather than respond to an already existing intrinsic motivation and that the interactions between intrinsic and extrinsic motivation are not sufficiently understood in the health care context [110, 113-115]. Furthermore, even in cases where quality improvement has been attributed to a specific P4P program, it remains unclear whether this was caused by the program itself or parallel developments. On top of that, it may be questioned whether a P4P program is the smartest way to enhance quality. To date, it is not well understood how this instrument compares to alternatives which come with a lower price tag. Tools such as public reporting and better hospital management and leadership are also positively associated with quality improvements and may represent a suitable alternative to P4P [116-119]. Finally, it is debated how the potential positive effects of P4P programs, such as greater attention to quality as a whole, weight against potential increases in health disparities resulting from factors such as adverse selection and an unreasonable penalization of hospitals caring for disadvantaged patient groups.

When criticizing P4P, however, three aspects should be put into consideration. To begin with, P4P programs are seldom evaluated. Investigations of programs such as the

Premier HQID, Advancing Quality and the HVBP to an increasing degree offer valuable information to policy makers on how to design a P4P. This, though, is Anglo-Saxon-dominated. Evaluations of programs beyond Canada, England and the United States, which go beyond descriptive statistics, operate with a longer time frame and allow for comparisons with a control group, are virtually non-existing. Second, evidence is often transferred from the ambulatory care sector to the inpatient sector without taking into account the different setting. For example, evidence on the crowding-out of motivation has been generated on the basis of individuals [110, 113, 114]. However, it has not been tested on complex entities, such as a hospital. As hospitals operate under different payment systems and are much more complex entities with a different role profile, evidence from individual physicians in ambulatory care should be viewed with utmost caution and used as an indication at best [120]. Third, all programs demonstrate methodological flaws. For example, they only reward top performers, or grant negligible rewards which are too low to effectively change investment patterns [89]. None of the programs in this review incorporate the manifold lessons learned researchers have identified over the last years.

In order to explore the full potential P4P holds in store and to minimize negative side-effects, policy makers are encouraged to take the following aspects into account. First, policy makers should formulate a clear overarching aim when introducing a P4P program which depends largely on the context of their country. If selected conditions suffer from poor quality, such as AMI and C-sections in South Korea, a P4P program which targets those conditions explicitly appears to be reasonable. If the quality of care as a whole is of concern, or if the P4P program shall be used to correct for deficiencies in the remuneration system as such, a broader approach would be suitable as currently applied in Luxemburg. This should be viewed in the context of the entire inpatient system structure. For example, a P4P program which targets waiting times or lengths of stay might be applicable for a system which remunerates on budgets, but not for a DRG-system. Hence, a P4P program can correct the deficiencies of a system which has generally been understood as suitable. If the P4P program shall be based on conditions,

emergency conditions are particularly suitable because hospitals have less influence on patient choice. Thus, potential unintended consequences of adverse selection are less likely to occur compared to a P4P program which targets elective conditions. To further minimize the potential impact of unintended consequences, it is recommended to introduce the P4P program as a pilot project in a selection of hospitals first before it is rolled out to the entire country. This allows policy makers to correct deficiencies in the P4P program before it affects all hospitals.

Second, policy makers are encouraged to formulate clear target levels depending on the P4P's aim. A P4P program stipulates that there is a certain 'gold standard' or desirable level of high quality of care. This requires a joint understanding between policy makers, providers and sickness funds alike on how this ought to be defined and how it can be measured. Linked with that, it requires all actors to define which quality of care ought to be understood as 'normal'. On the one hand, high quality of care can be understood as 'normal' and as the integer role of hospitals. If so, it appears odd to reward hospitals for what they should be doing in the first place and in this case, a system based on penalties might be more appropriate. On the other, the aims stipulated in the P4P program can be understood as outstanding quality ranging beyond 'normal'. In this case, it is sensible to reward hospitals which distinguish themselves from their compatriots by superior quality and additional effort.

Third, policy makers have to choose indicators accordingly. This requires a clear understanding on the structural prerequisites and the accompanying processes which determine good quality of care and on the outcomes which are desired. P4P programs in their very core serve the purpose of improving patient outcomes. Unfortunately, the assumption that meeting process indicators automatically improves outcomes does not hold true in all cases [121, 122]. Hence, including outcome indicators can correct for some insecurities in process indicators. It has to be taken into account that the inclusion of outcome indicators comes with the prerequisite of a comprehensive dataset attached. Due to that, policy makers have to decide on the data based on which hospitals are evaluated and rewarded. Information should be as objective, neutral and

comprehensive as possible. Administrative data particularly with information before and after the hospital admission should be given preference over self-reported data as the latter is clearly prone to fraud which might be further supported if awards are conditional on performance drawn from such data. Policy makers are encouraged to make the best possible data available to ensure the fairness and hence acceptance by providers. If the data set does not allow adjusting for environmental and patient characteristics, a hospital which is located in a disadvantaged area may be unreasonably punished for aspects beyond its control, whereas it may actually deserve to be compensated for its challenging environment. Hence, providers may judge the system to be unfair and try to circumvent it [123]. A comprehensive and objective dataset, though, allows for risk-adjustment and ensures a fair allocation of awards to help hospitals to improve their quality within their means.

Fourth, the design of the financial incentives should be thoughtfully aligned with the underlying understanding of policy makers on the P4P program. Policy makers can choose from a vast array of policy options and intervention points and all payment systems have their advantages and disadvantages [123, 124]. They are tasked with the challenging job to balance out budget constraints, effect maximization and acceptance by providers while taking specific problems faced by a country or region into account [124, 125]. Beyond that, the design of the incentive structure is a normative question which depends on how the purpose of the P4P system is understood and what is defined as normal. If the level of quality defined in a given P4P program is understood to be normal, it might appear odd to reward hospitals for attaining a level of quality they should be providing under normal circumstances anyway. On the other hand, a bonus payment can be understood to fill the financial gap that arises from providing higher quality care [89]. Indeed, there is a growing body of academic literature supporting the positive relationship between financial resources and quality of care [126]. Conditions, for which this link is particularly well understood, are those which are included in P4P programs such as the Premier HQID, the HVBP and Advancing Quality, namely AMI, pneumonia, congestive heart failure and to some extend also stroke [127].

In this case, paying an additional amount to compensate for costs arising from the provision of higher care seems sensible.

Over the design process, policy makers should take four aspects into account: To begin with, policy makers have to decide between an absolute and a relative score. With regards to that, providers appear to prefer a stepped absolute score over relative ranking [43, 125]. The former is considered to be more transparent and is associated with less uncertainty in terms of revenue [123]. Simple absolute scores are prone to a ceiling-effect whereby hospitals may have an incentive to improve their performance up to a performance threshold but not beyond [59, 124]. If simple absolute scores are used, they should be updated at regular intervals, for example annually, to ensure continual quality improvement.

Following from that, policy makers can choose between top-performer scores, attainment and improvement scores. Currently, the trend in P4P program design appears to be towards a blend of attainment and improvement scores with top-performer scores [60, 124]. In a qualitative study by the RAND corporation, providers unanimously welcomed the blending of such awards, which takes improvements over time into account [125]. Top-performer awards alone do not give any incentive to poor-performing hospitals, which are unlikely ever to meet the criteria. Indeed, such awards disproportionally favor already well-performing hospitals or hospitals with healthier patient populations [83, 123]. To address this shortcoming, more recent P4P programs have included attainment and improvement awards in their reward system, and programs that have already been implemented have changed their incentive structure in a similar fashion. For example, the HVBP included attainment and improvement scores from the start. Advancing Quality and the Premier HQID program have shifted their award structure towards a combination of top-performer and attainment awards [58, 73]. In the case of Premier HQID, this results in a fairer allocation of awards [73].

In addition, policy makers have to decide on whether to use sticks, carrots or both. When choosing among these approaches, policy makers must find a balance between budget constraints, political acceptance by hospitals and the magnitude of the effect. Programs with negative incentives, whether these be withholding payments, penalties alone, or a combination of bonuses and penalties, are often associated with greater effects [123]. Werner et al. showed in a simulation that if the bonus pool is held constant, a combination of bonuses and penalties can strengthen incentives if the money collected from penalties can be redistributed to the best performers as bonuses [124]. Quality improvements correlate positively with the size of incentives, this strategy correlated with the greatest improvements by hospitals [124, 128]. Withholding payments has also been associated with changes in hospital behavior that are greater than those achieved through bonus payments alone [123, 125]. Unfortunately, it is not yet clear how withholding payments compares to a combination of bonuses and penalties. It may come as no surprise that providers largely favor bonus payments [43, 125]. Programs that redistribute "old" money can be perceived as unfair compared to a heterogeneous distribution of bonuses [129]. In the former case, political opposition from providers can be expected to be greater, and compliance may be lower [123]. Hence, there is a trade-off between the magnitude of the effect and the budget constraints, on the one side, and the acceptance of the program on the other.

On top of that, they have to decide on the amount. A higher amount is understood to lead to greater effects [88]. In line with that, the low incentive payments provided within P4P programs have been criticized by providers. In England, for example, providers reported that the financial incentives offered by Best Practice Tariffs were negligible and did not play a role in hospitals' budget calculations [89]. Similarly, in Canada, it was argued that incentives amounting to less than 1% of a hospital's budget were too small to lead to changes in the emergency departments, such as the hiring of additional staff [90]. The Advancing Quality program in England initially set its awards at 4% of a hospital's budget in response to the 2% award in the Premier HQID, which

had been criticized as too low and cited as a reason for the mixed success of the program [61]. Unfortunately, the incentive structure of Advancing Quality changed after 6 months from a program that operated with bonuses of up to 4% to a system that applied withhold payments only and no longer granted bonuses. This makes it difficult to determine whether the initial 4% award had an effect on the behavior of participating hospitals. In total, policy makers are encouraged to make incentive payments of 4% or more available.

Fifth, policy makers should consider the involvement of interest groups in the crafting of the P4P program. This might foster the adaptation by hospitals and physicians because they have been integrated in the decision process and do not have the impression that the P4P program is not yet another policy which has been imposed in them by policy makers in a top-down approach [123, 130, 131]. In addition to that, medical societies might make valuable contributions on the selection of indicators to best measure the quality of care of a hospital [125].

Sixth, policy makers have to ensure that their P4P program is accompanied by an evaluation process which meets academic standards. To date, the possibility to draw conclusions from P4P programs is impeded by the absence of solid evaluations. Selected countries, such as the United Kingdom, have institutionalized the evaluation of programs. This should be regarded as self-evident given the financial volume which is at stake even if it is small in comparison to the total budget of acute cares. Furthermore, evaluations can point at deficiencies in the design and unveil unintended consequences through which a P4P program might do more harm than good. In addition, the criticism with regards to the lack of evaluations is not going to change if policy makers do not commission evaluations or support them by providing data. This includes that they have to accept the potential failure of a P4P program if the evaluation shows that the program could not live up to the expectations.

This paper has several important limitations. First, it suffers from a selection bias. Programs that were evaluated had been investigated in greater detail. Promising or otherwise interesting programs that had not undergone evaluation may have been ignored. Additionally, language limitations led to Eastern European countries not being investigated with the intensity they deserve. Second, this paper could not assess further vital aspects of the success of P4P programs. For example, it did not report information about the hospital level that had been incentivized. This could be the hospital as a whole, a hospital department, a clinical team or individual clinicians. This dimension had to be excluded because information was not available. Third, the evidence base for P4P programs remains weak. Thus, policy recommendations should be made with caution.

Additionally, several shortcomings make it difficult to identify casual effects. First, the limitations of the majority of studies only allow for preliminary conclusions to be drawn on the effect of P4P programs on quality of care [91]. Second, P4P programs are generally part of a larger package of policy changes. In Australia and Canada, for example, they are introduced in conjunction with changes in the remuneration system. In almost all cases, the P4P programs identified in our research were combined with public reporting. These alone may already lead to an improvement in quality of care. Several studies have indicated that public reporting has a positive effect on the quality improvement of hospitals [132]. Third, various programs change frequently. The HVBP, for example, changes the indicators, their weighting for the total performance score and the amount of incentives on an annual basis. Even if we take into account that hospitals are very price-sensitive, they might still require some time to adapt to the new scheme.

These difficulties, which hamper policy makers and researchers alike, call for additional research to better understand how the various design elements and intervention points of P4P are received by hospitals. To begin with, there is a lack of understanding which dynamics a P4P program unfolds within a hospital [120]. For example, it may be

that hospitals start reviewing the performance of clinical teams and departments or whether they rearrange the way they deliver care [89]. Additionally, while we can generally assume that greater financial resources improve quality of care, we do not know how much precisely hospitals need to arrive at better levels. As P4P intends to change investment decisions and to provide hospitals with the financial means they need to provide better quality of care, it is necessary to ensure that the additional money granted is sufficient to fill this gap [5, 133]. At the same time, it should not be too much in order not to waste money that could be better invested elsewhere. Furthermore, we do not know sufficiently enough about how P4P compares to alternative policy instruments to enhance quality of care. In times of financial constraints, policy makers may want to maximize the utility of public money. Policy makers will have to contrast P4P to public reporting, mandatory structural indicators for hospitals to perform certain interventions, such as minimum volumes, or a combination of them.

References

1. Witter, S., et al., *Paying for performance to improve the delivery of health interventions in low- and middle-income countries.* Cochrane Database Syst Rev, 2012. **2**: p. CD007899.
2. Lazear, E.P., *Performance Pay and Productivity.* American Economic Review, 2000. **90**(5): p. 1346-1361.
3. Prendergast, C., *The Provision of Incentives in Firms.* Journal of Economic Literature, 1999. **37**(1): p. 7-63.
4. OECD, *Performance-related pay policies for government employees.* OECD Publishing, 2005.
5. Rosenthal, M.B., et al., *Paying for quality: providers' incentives for quality improvement.* Health Aff (Millwood), 2004. **23**(2): p. 127-41.
6. Rosenthal, M.B., et al., *Early experience with pay-for-performance: From concept to practice.* JAMA, 2005. **294**(14): p. 1788-1793.
7. Eijkenaar, F., *Pay for performance in health care: an international overview of initiatives.* Med Care Res Rev, 2012. **69**(3): p. 251-76.
8. Campbell, S., et al., *Quality of Primary Care in England with the Introduction of Pay for Performance.* New England Journal of Medicine, 2007. **357**(2): p. 181-190.
9. Mehrotra, A., et al., *Pay for performance in the hospital setting: what is the state of the evidence?* Am J Med Qual, 2009. **24**(1): p. 19-28.
10. Minister van Sociale Zaken en Volksgezondheid, *Plan van aanpak hervorming ziekenhuisfinanciering.* 2015.
11. Deutscher Bundestag, *Gesetzesentwurf der Fraktionen der CDU/CSU und SPD. Entwurf eines Gesetzes zur Reform der Strukturen der Krankenhausversorgung (Krankenhausstrukturgesetz - KHSG).* Drucksache 18/5372, 2015.
12. Campbell, S.M., et al., *Effects of Pay for Performance on the Quality of Primary Care in England.* New England Journal of Medicine, 2009. **361**(4): p. 368-378.
13. OECD/WHO, *Paying for Performance in Health Care*. 2014: Open University Press - McGraw-Hil.
14. Ryan, A. and J. Blustein, *Making the Best of Hospital Pay for Performance.* New England Journal of Medicine, 2012. **366**(17): p. 1557-1559.
15. OECD. *Health Systems Characteristics Survey 2012, and Secretariat's estimates*. 2014; Available from: http://www.oecd.org/els/health-systems/paying-providers.htm.
16. Gosden, T., et al., *Capitation, salary, fee-for-service and mixed systems of payment: effects on the behaviour of primary care physicians.* Cochrane Database Syst Rev, 2000(3): p. CD002215.
17. O'Reilly, J., et al., *Paying for hospital care: the experience with implementing activity-based funding in five European countries.* Health Economics, Policy and Law, 2012. **7**(Special Issue 01): p. 73-101.
18. Marmor, T., R. Freeman, and K. Okma, *Comparative perspectives and policy learning in the world of health care.* Journal of comparative policy analysis: Research and practice, 2005. **7**(4).
19. Dolowitz, D. and D. Marsh, *Who Learns What from Whom: a Review of the Policy Transfer Literature.* Political Studies, 1996. **44**(2): p. 343-357.
20. Donabedian, A., *Evaluating the quality of medical care.* Milbank Mem Fund Q, 1966. **44**(3): p. Suppl:166-206.
21. AIHW, *Australian hospital statistics 2013-14. Emergency department care*. 2014: Canberra.
22. AIHW, *Australian hospital statistics 2013-14. Elective surgery waiting times*. 2014: Canberra.
23. Duckett, S., et al., *Pay for performance in Australia: Queensland's new Clinical Practice Improvement Payment.* J Health Serv Res Policy, 2008. **13**(3): p. 174-7.
24. Stockwell, A., *Evaluation of financial incentives as a quality improvement strategy in the public hospital context: Clinician attitudes, design variables, and economic costs*. 2010, Queensland University of Technology: Brisbane.

25. Cheng, A.H. and J.M. Sutherland, *British Columbia's pay-for-performance experiment: part of the solution to reduce emergency department crowding?* Health Policy, 2013. **113**(1-2): p. 86-92.
26. Cheng, I., et al., *Implementing wait-time reductions under Ontario government benchmarks (Pay-for-Results): a Cluster Randomized Trial of the Effect of a Physician-Nurse Supplementary Triage Assistance team (MDRNSTAT) on emergency department patient wait times.* BMC Emerg Med, 2013. **13**: p. 17.
27. Vermeulen, M.J., et al., *The Effect of Pay for Performance in the Emergency Department on Patient Waiting Times and Quality of Care in Ontario, Canada: A Difference-in-Differences Analysis.* Annals of Emergency Medicine, 2015.
28. Vermeulen, M.J., et al., *Evaluation of an emergency department lean process improvement program to reduce length of stay.* Ann Emerg Med, 2014. **64**(5): p. 427-38.
29. Kristensen, S.R., M. Bech, and J.T. Lauridsen, *Who to pay for performance? The choice of organisational level for hospital performance incentives.* Eur J Health Econ, 2015.
30. Ministère des affaires sociales, d.l.s.e.d.d.d.f., *Arrêté du 31 mars 2015 fixant la liste des établissements de santé éligibles à un financement complémentaire portant sur l'amélioration de la qualité et de la sécurité des soins en 2015,.* Journal officiel de la Républicque française, 2015. **83**.
31. HAS, *Présentation du modèle d'Incitation Financière à l'Amélioration de la Qualité (IFAQ), Point d'étape.* 2013, Saint-Denis La Plaine.
32. Ministère des affaires sociales, d.l.s.e.d.d.d.f., *Circulaire No. DGOS/R1/2014/366 du 29 décembre 2014 relative à la campagne tarifaire 2014 des établssements de santé,*. 2014, Paris.
33. Weiss-Meilik, A. and G. Uretzky, *The effect of pay for performance (P4P) on quality of care in cardiac surgery.* NIHP Publications, 2012.
34. Israeli Ministry of Health, *יקרו בל אירועי :נושא [Topic: "Never-events"]. Circular by the Director General.* 2011.
35. Colais, P., et al., *The impact of a pay-for-performance system on timing to hip fracture surgery: experience from the Lazio Region (Italy).* BMC Health Serv Res, 2013. **13**: p. 393.
36. Pinnarelli, L., et al., *What drives hospital performance? The impact of comparative outcome evaluation of patients admitted for hip fracture in two Italian regions.* BMJ Quality & Safety, 2012. **21**(2): p. 127-134.
37. Inoue, Y., S. Jeong, and K. Konto, *Changes in hospital performance after implementation of pay-for-performance in Japan.* Clinical Audit, 2011. **3**: p. 17-23.
38. Jeong, S.K., Katsunori; Shiraishi, Nariaki; Inoue, Yusuke, *An evaluation of the quality of post-stroke rehabilitation in Japan.* Clinical Audit 2010. **2**: p. 59-66.
39. 鄭丞媛, 井., *質に基づく支払い（Pay for performance:P4P）の動向と今後のあり方 [Quality-based payments (Pay-for-performance): Trends and future directions].* Quaterly social security research. **48**(2): p. 186-196.
40. FHL-CNS, C.d.e., *Module 2014 - Incitant qualite*. 2013: Bertrange.
41. CNS (formerly UCM), EHL, and TU, *Convention entre l'Union des caisses de maladie et l'Entente des hôpitaux lugembourgeois, conclue en exécution des articles 74 et suivants du Code des assurances sociales.* 1995.
42. CNS and FHL, *Convention entre Caisse nationale de santé et la Fédération des hôpitaux luxembourgeois, cenclue en exécution de l'article 74 et suivants du Code de la sécurité sociale.* Mémorial A-030 du 25 février 2013, 2013: p. 487-516.
43. Vitali, S. *EFQM et le modèle des incitants qualité. 15 anées de démarches communes EHL-CNS*. in *Quelles méthodes pour améliorer la performance clinique?* 2012. Namur.
44. Kiviluoto, L. and C. Beck Olsen, *Kvalitetsbasert finansiering 2015*. 2014, Oslo: Helsedirektoratet.
45. Jeong, H.-S., *Designing an Effective Pay-for-performance System in the Korean National Health Insurance.* Journal of Preventive Medicine and Public Health, 2012. **45**(3): p. 127-136.

46. HIRA, *가감지급시범사업 결과 보고서 [Report on results of a pay-for-performance pilot project].* 2011.
47. Lee, J.Y., et al., *Healthcare organizations' attitudes toward pay-for-performance in Korea.* Health Policy, 2012. **108**(2-3): p. 277-85.
48. Jeong, H.S., *Designing an effective pay-for-performance system in the Korean National Health Insurance.* J Prev Med Public Health, 2012. **45**(3): p. 127-36.
49. Högberg, M. and S. Lindvall, *Ersättningssystem inom hälso- och sjukvård. Erfarenheter, utvecklingslinjer och vägval för den mer specialiserade vården (Sjukhus, slutenvård och öppenvård)*. 2012, Stockholm: Forum för health policy.
50. Jacobsson, F. and S. Lindvall, *Utveckling av ersättningssystem inom hälso- och sjukvård. En översikt av tillämpning, erfarenheter och utvecklingslinjer*. 2007, Stockholm: Sveriges Kommuner och Landsting.
51. Västra Götalandsregionen, *Regiongemensamma ersättningsmodeller för sjukhusvård. Uppföljning och förtydligande*. 2010, Vänersborg.
52. Landstinget i Uppsala län, *Uppföljning av målrelaterad ersättning 2010*. 2011, Uppsala.
53. Yuzden, G.E. and J. Yildirim, *A Qualitative Evaluation of the Performance-based Supplementary Payment System in Turkey: Physicians' Perspectives.* Journal of Health Management, 2014. **16**(2): p. 259-270.
54. Gok, M. and E. Altındağ, *Analysis of the cost and efficiency relationship: experience in the Turkish pay for performance system.* The European Journal of Health Economics, 2015. **16**(5): p. 459-469.
55. Kaptanoğlu, A.Y., *İstanbul'da Kamu Hastanelerinde Performans Esaslı Geri Ödeme Sistemleri/ Performance Based Supplementary Payment Systems in Istanbul Public Hospitals.* Yükseköğretim ve Bilim Dergisi/Journal of Higher Education and Science. **3**(2): p. 128-132.
56. Erkan, A., *Performansa Dayalı Ödeme: Sağlık Bakanlığı Uygulaması.* Maliye Dergisi, 2011. **160**(423-438).
57. Sülkü, S.N., *Performansa Dayalı Ek Ödeme Sisteminin Kamu Hastanelerinin Verimliliği Üzerine Etkileri.* Maliye Dergisi, 2011. **160**: p. 242-268.
58. Kristensen, S.R., R. McDonald, and M. Sutton, *Should pay-for-performance schemes be locally designed? evidence from the commissioning for quality and innovation (CQUIN) framework.* Journal of Health Services Research & Policy, 2013. **18**(2 suppl): p. 38-49.
59. Kamath, P., et al., *Improving outcomes in asthma by Commissioning Quality and Innovation (CQUIN).* Archives of Disease in Childhood, 2012. **97**(Suppl 1): p. A27-A28.
60. Khan, S.K., et al., *Achieving best practice tariff may not reflect improved survival after hip fracture treatment.* Clinical Interventions in Aging, 2014. **9**: p. 2097-2102.
61. Audit Commission, *Best practice tariffs and their impact*. 2012: London.
62. Kristensen, S.R., et al., *Long-term effect of hospital pay for performance on mortality in England.* N Engl J Med, 2014. **371**(6): p. 540-8.
63. McDonald, R., et al., *A qualitative and quantitative evaluation of the introduction of best practice tariffs. An evaluation report commissioned by the Department of Health.* 2012.
64. McDonald, R., et al., *A quallitative and quantitative evaluation of the Advancing Quality pay-for-performance programme in the NHS North West.* Health Services Delivery Research, 2015. **3**(23).
65. Sutton, M., et al., *Reduced mortality with hospital pay for performance in England.* N Engl J Med, 2012. **367**(19): p. 1821-8.
66. McDonald, R., et al., *Evaluation of the Commissioning for Quality and Innovation Framework*. 2013, London.
67. Berthiaume, J., et al., *Aligning financial incentives with quality of care in the hospital setting.* Journal for healthcare quality, 2006. **28**(2): p. 36-44, 51.
68. Ashby, J., et al., *The Relationship of Hospital Quality and Cost per Case in Hawaii.* INQUIRY: The Journal of Health Care Organization, Provision, and Financing, 2012. **49**(1): p. 65-74.
69. Berthiaume, J.T., et al., *Aligning financial incentives with "Get With The Guidelines" to improve cardiovascular care.* Am J Manag Care, 2004. **10**(7 Pt 2): p. 501-4.

70. Waters, T.M., et al., *Effect of medicare's nonpayment for hospital-acquired conditions: Lessons for future policy.* JAMA Internal Medicine, 2015. **175**(3): p. 347-354.
71. Lee, G.M., et al., *Effect of Nonpayment for Preventable Infections in U.S. Hospitals.* New England Journal of Medicine, 2012. **367**(15): p. 1428-1437.
72. McNair, P.D., H.S. Luft, and A.B. Bindman, *Medicare's Policy Not To Pay For Treating Hospital-Acquired Conditions: The Impact.* Health Affairs, 2009. **28**(5): p. 1485-1493.
73. Lindenauer, P.K., et al., *Public reporting and pay for performance in hospital quality improvement.* N Engl J Med, 2007. **356**(5): p. 486-96.
74. Lindenauer, P.K., *Public reporting and pay-for-performance programs in perioperative medicine: are they meeting their goals?* Cleve Clin J Med, 2009. **76 Suppl 4**: p. S3-8.
75. Ryan, A.M., *Effects of the Premier Hospital Quality Incentive Demonstration on Medicare patient mortality and cost.* Health Serv Res, 2009. **44**(3): p. 821-42.
76. Ryan, A.M., et al., *The early effects of Medicare's mandatory hospital pay-for-performance program.* Health Serv Res, 2015. **50**(1): p. 81-97.
77. Ryan, A.M., et al., *The effect of Phase 2 of the Premier Hospital Quality Incentive Demonstration on incentive payments to hospitals caring for disadvantaged patients.* Health Serv Res, 2012. **47**(4): p. 1418-36.
78. McIlvennan, C.K., Z.J. Eapen, and L.A. Allen, *Hospital Readmissions Reduction Program.* Circulation, 2015. **131**(20): p. 1796-1803.
79. Rajaram, R., et al., *Hospital characteristics associated with penalties in the centers for medicare & medicaid services hospital-acquired condition reduction program.* JAMA, 2015. **314**(4): p. 375-383.
80. CMS, *Terms and Conditions between Premier Inc. and CMS of the Hospital Quality Incentive Demonstration. First Amendment.* 2007.
81. CMS, *Terms and Conditions. Premier, Inc. Hospital Quality Incentive Demonstration.* 2006.
82. Colais, P., et al., *The impact of a pay-for-performance system on timing to hip fracture surgery: experience from the Lazio Region (Italy).* BMC Health Services Research, 2013. **13**(1): p. 1-7.
83. Kim, S.M., et al., *Korean National Health Insurance value incentive program: achievements and future directions.* J Prev Med Public Health, 2012. **45**(3): p. 148-55.
84. AQuA/NHS Northwest. *Conditions*. [cited 2015 30.07.]; Available from: http://www.advancingqualitynw.nhs.uk/measures/.
85. US Senate, *Patient Protection and Affordable Care Act. Section 3001. Hospital value-based purchasing program*. 2009: Washington D.C.
86. Ryan, A.M., et al., *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): p. 1418-1436.
87. Tompkins, C.P., A.R. Higgins, and G.A. Ritter, *Measuring Outcomes And Efficiency In Medicare Value-Based Purchasing.* Health Affairs, 2009. **28**(2): p. w251-w261.
88. Werner, R.M. and R.A. Dudley, *Making The 'Pay' Matter In Pay-For-Performance: Implications For Payment Strategies.* Health Affairs, 2009. **28**(5): p. 1498-1508.

The Hamburg Center for Health Economics is a joint center of Universität Hamburg and the University Medical Center Hamburg-Eppendorf (UKE).

hche | Hamburg Center for Health Economics

Esplanade 36
20354 Hamburg
Germany
Tel: +49 (0) 42838-9515/9516
Fax: +49 (0) 42838-8043
Email: info@hche.de
http://www.hche.de
ISSN 2191-6233 (Print)
ISSN 2192-2519 (Internet)

HCHE Research Papers are indexed in RePEc and SSRN.
Papers can be downloaded free of charge from http://www.hche.de.