Health policy 164 (2026) 105365

Contents lists available at ScienceDirect

Health policy

journal homepage: www.elsevier.com/locate/healthpol

Does pay for performance affect socioeconomic inequalities in access? Evidence from hospital specialised care in England

Alberto Núñez-Elvira [a], Yan Feng [b], Søren Rud Kristensen [a,c], Paula Lorgelly [d], Rachel Meacock [e], Luigi Siciliani [f,*], Matt Sutton [e]

[a] Institute of Global Health Innovation, Imperial College London, London, United Kingdom
[b] Health Economics and Policy Research Unit, Centre for Evaluation and Methods, Wolfson Institute of Population Health, Queen Mary University of London, London, United Kingdom
[c] Danish Centre for Health Economics (DaCHE), Department of Public Health, University of Southern Denmark, Odense, Denmark
[d] School of Population Health & Department of Economics, University of Auckland, New Zealand
[e] Health Organisation, Policy and Economics (HOPE) Research Group, School of Health Sciences, The University of Manchester, Manchester, United Kingdom
[f] Department of Economics and Related Studies, University of York, York, United Kingdom

ARTICLE INFO

Keywords:
Pay for performance
Quality
Efficiency
Hospitals
Socioeconomic inequalities

ABSTRACT

Pay for performance aims to improve quality and efficiency in the health sector but may widen inequalities. We investigate how pay for performance for specialised hospital care in England affected socioeconomic inequalities in access. We focus on two clinical areas: trauma care aimed at reducing delayed discharges from adult critical care; and internal medicine aimed at reducing in-hospital waiting time and length of stay for patients requiring urgent coronary bypass grafting. Both were part of the Prescribed Specialised Services Commissioning for Quality and Innovation. Using patient-level administrative data from Hospital Episodes Statistics in 2012/13–2016/17, we employ difference-in-difference models to estimate the impact of these schemes across socioeconomic status. Our treatment group comprises hospitals that adopted the scheme, and our control group the remaining eligible hospitals. For trauma care, we measure the impact of the scheme on discharge delays and the probability of an overnight discharge. For urgent coronary bypass, we measure pre-surgery inpatient waiting time, length of stay, 30-day and one-year mortality, and hospital-acquired infections. For trauma care we find the scheme widened inequalities by reducing delays that favoured more patients in the least income-deprived quintile (by 2.4 h or 30.4 % at the sample mean) than in the most income-deprived quintile (by 1.3 h). We find no effect or socioeconomic differences across outcomes for patients requiring an urgent coronary bypass.

Research in context

(1) What is already known about the topic?

Pay-for-performance schemes are varied and can incentivise process measures of quality, health outcomes or activities that reduce costs. The evidence on P4P is mixed and context dependent.

(2) What does this study add to the literature?

Our study contributes to the literature on pay for performance by examining whether there is a trade-off between improving performance and increasing socioeconomic inequalities. We contribute to the limited but growing literature investigating the heterogenous effects of pay for performance by socioeconomic status empirically. We also contribute to the literature on differences in access by socioeconomic status on different areas of care.

(3) What are the policy implications?

There is scope to further refine the design of pay-for-performance schemes that include equity considerations. There may be trade-offs between improving performance and inequalities of access to care. It is important that evaluations of pay-for-performance schemes assess more systematically the effect, not only on the average performance but also across different socioeconomic groups.

* Corresponding author.
E-mail address: luigi.siciliani@york.ac.uk (L. Siciliani).

https://doi.org/10.1016/j.healthpol.2025.105365
Received 17 December 2024; Accepted 28 May 2025
Available online 23 June 2025

1. Background

Policymakers aim at improving quality and efficiency in the health sector. A popular policy lever is the introduction of pay for performance (P4P) schemes for secondary and primary care. P4P schemes build on payment systems that traditionally focus on throughput by giving direct financial incentives to improve quality. P4P schemes can incentivise process measures of quality, health outcomes (mortality and readmissions rates) or activities that reduce costs, such as same-day discharges. The evidence is mixed and context dependent [14,15,24].

Reduction in health inequalities is also a pervasive policy objective, often with a focus on socioeconomic inequalities in health ([10]; Health and Care Act, 2022). Inequalities in quality of care can be a source of health inequalities. One concern with P4P is that improvements in provider performance may come at the cost of increased inequalities [1,6,19]. This is because it may cost less to treat patients with higher socioeconomic status to achieve a certain performance.

This study tests whether P4P increases socioeconomic inequalities in quality of care. We focus on a P4P scheme for hospitals in the National Health Service (NHS) in England which focussed on improving performance on specialised care for patients with complex conditions, known as the Prescribed Specialised Services Commissioning for Quality and Innovation (PSS CQUIN) [7]. This national P4P scheme aimed to improve quality of specialised services and achieving value for money. Between 2016–19 the potential incentive payments nationally were approximately £900 million. The incentive scheme was developed centrally with payment based on performance targets. It covered a range of clinical areas, such as cancer, internal medicine, trauma, women and children's health and mental health. Feng et al. [8] found that the scheme had limited effect. Across all areas, performance improved only in the clinical area of trauma.

We focus on two clinical areas of specialised care. The first relates to trauma care and aimed at reducing delayed discharges from adult critical care. The rationale is that patients experiencing delays in planned discharge from critical to lower-level care are more likely to experience a nighttime discharge or an expedited discharge to accommodate another patient, and these expedited discharges can lead to reduced patient experience characterised by unnecessary additional hand-offs of care and the inherent risks this poses. The second relates to internal medicine and aimed at reducing in-hospital waiting times and length of stay for patients requiring urgent coronary artery bypass grafting (CABG) to reduce risk of mortality and hospital-acquired infections. In-hospital waiting times differ from inpatient waiting times because they refer to the time for a surgery that patients experience once already admitted at the hospital (rather than before hospital admission). We focus on these two schemes because they are concerned with improving patient flows within the hospital captured by different measures of length of stay. Moreover, we know from a previous evaluation [8] that performance improved for trauma, and this gives an opportunity to test the heterogenous effect of the P4P scheme across socioeconomic status. The scheme on internal medicine, which focuses on CABG, is clinically important and covers a comprehensive range of outcomes. It therefore provides a means to test if the previous null finding hides heterogeneity in the effect across socioeconomic status.

Using patient-level administrative data from Hospital Episodes Statistics covering 2012/13–2016/17, we employ difference-in-difference models to estimate the impact of this scheme across patients with different socioeconomic status. Our treatment group comprises hospitals that adopted the scheme, and we use the remaining hospitals that were eligible for the scheme but did not take it up as our control group. Although the scheme was voluntary, we use propensity score matching based on hospital characteristics to minimise selection bias. Socioeconomic status is measured by income deprivation measured at small geographical area level (Lower Super Output Area, LSOA) following previous studies [12,17]. We find that the effect of the P4P scheme in trauma was more pronounced for less deprived patients. We find no differences by deprivation for patients requiring an urgent coronary bypass.

Our study contributes to the literature on P4P by examining whether there is a trade-off between improving performance and increasing socioeconomic inequalities. Theoretical studies have highlighted that financial incentives can increase inequalities in access in relation to tariff adjustment [16], hospital competition [20] and reimbursement systems [9]. We contribute to the limited literature investigating heterogenous effects of P4P by socioeconomic status empirically [1,6] where research focuses mostly on primary care and less on secondary care [11]. We also contribute to the literature on differences in access by socioeconomic status ([22] for emergency care; [3,12] for cancer; [17,21] for elective care).

1.1. *Institutional setting*

NHS hospital care is primarily funded through general taxation and provided without out-of-pocket costs to UK residents. Hospitals provide elective care, urgent and emergency care, and specialised services. Patients need a referral from a general practitioner (GP) to access hospital treatment, except for accident and emergency (A&E). Most NHS hospital care is provided by public hospitals funded via a blended reimbursement system combining elements of activity-based payment and block contracts [23].

In 2009, a national P4P scheme known as the Commissioning for Quality and Innovation (CQUIN) framework was introduced. It linked hospitals' income to quality improvements (NHS [18]). Since 2012, the national CQUIN framework included two schemes: the Clinical Commissioning Group CQUIN (CCG CQUIN) scheme, which incentivises care commissioned locally by CCG purchasers; and Prescribed Specialised Services CQUIN (PSS CQUIN) schemes covering specialised services commissioned nationally by NHS England but managed locally by commissioning teams. It was launched in 2013 to improve the quality of specialised service and deliver value for money [7].

Eligible providers and their local commissioning hubs discussed what schemes each provider could participate in [7]. There were 26 schemes in 2016/17. We study two schemes: the "Adult Critical Care Timely Discharge" aimed at improving quality of Trauma services (TR1), and the "Reducing Cardiac Surgery Non-elective Inpatient Waiting" within the Internal Medicine programme (IM1).

TR1 aimed "To reduce delayed discharges from ACC [adult critical care] to ward level care in the same hospital by improving bed management in ward based care, thus removing delays and improving flow." IM1 aimed to "ensure that patients referred for Coronary Artery Bypass Grafting (CABG), semi urgently, have CABG as an inpatient (with or without transfer) within seven days of an angiogram (wherever that takes place) or within seven days of transfer to a non-elective pathway (whichever is the later)" (NHS [18]). They were introduced on the 1st April 2016 and lasted one financial year.

All providers with critical care beds could take up the TR1 scheme, and 63 out of 135 eligible providers did. For IM1, all providers that perform semi-urgent CABG were eligible to join the scheme. Out of 28 eligible providers, 12 providers participated in the scheme.

The incentive design differed. For TR1, a baseline incentive payment set relative to hospitals' proportion of critical care discharges within a given hospital's tier was reduced by fixed penalties per patient that were more than four (£325) or 24 h delayed (+£500). An alternative scheme with penalties of £650 and £1000 respectively was available to selected trusts.

IM1 attracted two payments: a payment up to £10,000 to cover setting-up costs and monitoring; and a bonus rewarding reductions in days waited by patients beyond seven days relative to an agreed target. Each targeted reduction in waiting beyond seven days was worth up to £150 per day. The actual payment was determined by the proportion of achieved reduction as a proportion of the targeted reduction. The contract value paid by NHS England was £10.8 million for TR1 and £1.6

Table 1
Descriptive statistics.

	TR1: Adult critical care timely discharge			IM1: Cardiac Surgery Non-elective In-hospital Waiting Time		
	Control group	Treatment group	Whole sample	Control group	Treatment group	Whole sample
N	132,880 (26.0 %)	378,463 (74.0 %)	511,343 (100.0 %)	3829 (32.9 %)	7824 (67.1 %)	11,653(100.0 %)
Discharge delay > 4 h	0.487 (0.500)	0.474 (0.499)	0.478 (0.499)			
Discharge delay (hours)	7.575 (12.414)	7.949 (13.662)	7.852 (13.350)			
Night discharge (20h-7 h)	0.200 (0.400)	0.176 (0.381)	0.182 (0.386)			
Waiting for CABG > 7 days				0.518 (0.500)	0.589 (0.492)	0.566 (0.496)
Wait for CABG (days)				9.417 (6.250)	10.016 (5.950)	9.819 (6.057)
Length of stay (days)				19.760(10.277)	19.202 (9.672)	19.385 (9.878)
Death within 30 days				0.031 (0.173)	0.025 (0.156)	0.027 (0.161)
Death within 365 days				0.058 (0.233)	0.054 (0.226)	0.055 (0.229)
Patient safety incident				0.137 (0.343)	0.129 (0.335)	0.131 (0.338)
IMD Income quintile						
Q1(most deprived)	26.0 %	25.5 %	25.6 %	20.0 %	24.3 %	22.9 %
Q2	19.9 %	22.2 %	21.6 %	20.7 %	21.2 %	21.0 %
Q3	18.4 %	18.8 %	18.7 %	22.7 %	19.8 %	20.8 %
Q4	18.7 %	17.6 %	17.9 %	19.0 %	19.2 %	19.1 %
Q5 (least deprived)	17.0 %	15.9 %	16.2 %	17.6 %	15.5 %	16.2 %
Female	44.0 %	42.5 %	42.9 %	19.7 %	21.5 %	20.9 %
Age-band						
18–30	5.8 %	6.5 %	5.8 %	0.2 %	0.1 %	0.1 %
31–40	5.8 %	6.5 %	5.8 %	0.7 %	0.7 %	0.7 %
41–50	10.4 %	10.5 %	10.4 %	7.2 %	6.6 %	6.8 %
51–60	15.9 %	16.5 %	15.9 %	20.2 %	20.9 %	20.7 %
61–70	24.1 %	24.4 %	24.1 %	33.7 %	32.3 %	32.8 %
71–80	24.6 %	24.2 %	24.6 %	31.4 %	31.7 %	31.6 %
81+	13.4 %	11.4 %	13.4 %	6.6 %	7.8 %	7.4 %
Grouped Charlson index						
No co-morbidities	25.1 %	24.9 %	25.1 %	12.0 %	9.3 %	10.2 %
One comorbidity	24.2 %	23.4 %	24.2 %	33.7 %	38.9 %	37.2 %
2 or more comorbidities	50.6 %	51.6 %	50.6 %	54.3 %	51.8 %	52.7 %
Year						
2012/13	15.8 %	17.2 %	15.8 %	18.3 %	19.5 %	19.1 %
2013/14	19.6 %	20.1 %	19.6 %	20.2 %	20.7 %	20.5 %
2014/15	20.5 %	19.1 %	20.5 %	21.5 %	18.3 %	19.4 %
2015/16	22.2 %	21.5 %	22.2 %	21.1 %	21.0 %	21.0 %
2016/17 (After)	22.0 %	22.2 %	22.0 %	18.9 %	20.5 %	20.0 %

million for IM1 [8]. In total, 2.5 % of each provider's specialised services contract value was linked to incentives.

2. Methods

2.1. Data

We use data from Hospital Episodes Statistics (HES) covering all NHS admissions in England, for financial years 2012/13 to 2016/17. For both schemes, 2016/17 was the year the incentive was applied. Our sample includes all patients admitted to hospitals eligible for both schemes. For TR1, all adult critical care discharges were included except patients who died during critical care or were discharged to other hospitals. For IM1, CABG procedures are defined by primary codes; angiography procedures by codes in the inpatient and outpatient setting; and transfers by date and method of admission. A patient is semi-urgent if the procedure takes place during a non-elective (emergency) inpatient admission 2–30 days after angiography or transfer from another hospital.

For TR1, our final sample consists of 511,343 patients discharged from Adult Critical Care. 378,463 patients are from 63 hospitals in the treatment group and 132,880 patients from 61 hospitals in the control group. For IM1, our final sample includes 11,653 semi-urgent CABG patients. 7824 (67.1 %) patients are from 12 hospitals in the treatment group and 3829 (32.9 %) patients are from 16 hospitals in the control group.

For TR1, we examine the impact on discharge delays measured both in hours and as a dummy variable taking the value 1 if the delay is longer than four hours and zero otherwise. We also examine the probability of night discharges using a dummy variable taking the value 1 if a discharge from adult critical care occurred during the night (between 8pm and 7am).

For IM1, our primary outcome is pre-surgery in-hospital waiting time: the number of days between i) the date the patient has an angiogram procedure or the date of admission for patients transferred from another provider, and ii) the date the CABG procedure is performed. Hospitals are accountable for a patient who is transferred to the non-elective pathway after having an angiogram. As the timing of the pathway switch is unobservable for transferred patients, for these patients we use the admission date at the receiving hospital. We analyse this outcome as a continuous variable and as a dummy variable equal to 1 if the patient waits longer than 7 days for surgery, and zero otherwise. We also examine the impact on length of stay, mortality and patient safety. Length of stay (LOS) is the time between admission for CABG procedure and discharge from hospital. Mortality is a dummy equal to 1 if death occurs within 30/365 days from discharge. Patient safety events include any hospital infection acquired by the patient: infections due to medical care (PSI7), post-operative events such as haemorrhage or haematoma (PSI9), physiologic and metabolic derangement (PSI10), pulmonary embolism or deep vein thrombosis (PSI12), sepsis (PSI13), and accidental puncture or laceration (PSI15). For both schemes, we control for sex, age (in 10-year bands) and the Charlson comorbidities index [4].

2.2. Regression model

To identify the heterogeneous effect of the policy across different socioeconomic groups, we estimate a difference-in-difference model (see Online Supplementary Material (OSM) Part A) where we regress our outcome of interest for patient i in hospital h in year t against the following variables: i) year fixed effects to allow for a time trend (e.g. improvements of health outcomes due to new technologies and other factors); ii) hospital fixed effects, which account for time-invariant unobserved hospital factors; iii) control variables at the patient level (e.g. 10-year age-band, sex, and sex-age-band interactions, Charlson index indicating if the individual had none, one, or two or more comorbidities); iv) our key variables of interest, which are the interactions of the post-policy period with the treatment group and socioeconomic status (split into quintiles, from least to most deprived using the Index of Multiple Deprivation, income domain): these coefficients identify the heterogenous effect of the policy in the treatment group post-policy across five different socioeconomic groups and show whether the policy had a larger effect amongst more or less deprived patients. Our treatment group is captured by a dummy variable equal to one if hospital h is in the treatment group, and equal to zero if in the control group. The post-policy period is a dummy variable equal to one for the post-policy period (2016/17) and zero in the pre-policy period (2012/13–2015/16). We allow the year time trend to differ between the treatment and the control group. Standard errors are clustered at the provider level. We also test if the estimated DiD coefficient for Q1 differs against each of the other quintiles.

One disadvantage of the baseline specification is that it does not provide information on the socioeconomic gradient before the policy. We therefore employ a second more flexible model, which includes i) socioeconomic status (again split into quintiles); ii) socioeconomic status interacted with the treatment group; iii) socioeconomic status interacted with the control group and the post-policy period; iv) socioeconomic status interacted with the treatment group and the post-policy period; and v) year and hospital fixed effects. The coefficients associated with these variables give i) the socioeconomic gradient in the control group in the pre-policy period; ii) the socioeconomic gradient in the treatment group in the pre-policy period; iii) the difference in the socioeconomic gradient in the control group in the post-policy period relative to the pre-policy period, and iv) the difference in the socioeconomic gradient in the treatment group in the post-policy period relative to the pre-policy period. We use this model to predict the gradient before and after the policy period both in the treatment and the control group using the margins command in Stata (v.18). Finally, we use an event-study approach to compare yearly differences in outcomes between treatment and control groups by socioeconomic status comparing Q1 differences to differences for Q2–5.

One potential issue is that the decision of joining the CQUIN scheme is voluntary, and therefore hospitals in the control group may be systematically different from the treatment group (if for example they decided not to join the scheme due to concerns of low quality). We address this possible concern through a propensity-score matching approach, where we regress the probability of joining the scheme against a set of hospital characteristics (similarly to [8]). Specifically, these were the total number of beds in the hospital, the proportion of doctors out of the total number of hospital staff, the Market Forces Factor index and binary indicators for whether a provider had Foundation Trust status, was a teaching hospital, or was located in London. We then match hospitals in the treatment and the control group based on a "nearest neighbour" approach without replacement. Balance tables are included in the OSM Part B. They show that the matching approach reduced the imbalance between the treatment and the control groups.

Table 2
Regression results. Adult critical care timely discharge (TR1).

	Discharge delay > 4 h	Discharge delay (hours)	Nightly discharge
IMD Income quintile # Treated # After			
Q1 # Treated # After	−0.059	−1.292*	0.028
	(0.032)	(0.589)	(0.014)
Q2 # Treated # After	−0.077*	−1.717**	0.012
	(0.030)	(0.544)	(0.014)
Q3 # Treated # After	−0.071*	−1.675**	0.003
	(0.031)	(0.599)	(0.015)
Q4 # Treated # After	−0.074*	−1.920**	−0.006
	(0.031)	(0.611)	(0.015)
Q5 # Treated # After	−0.086*	−2.389**	−0.005
	(0.034)	(0.604)	(0.014)
Year # Treated			
2012/13 # Treated	−0.038	−1.941**	−0.025
	(0.032)	(0.694)	(0.014)
2013/14 # Treated	−0.077	−1.996*	−0.011
	(0.041)	(0.976)	(0.012)
2014/15 # Treated	−0.051	−1.000	−0.007
	(0.037)	(0.761)	(0.008)
Year			
2012/13	0.020	0.497	−0.013
	(0.026)	(0.600)	(0.008)
2013/14	0.022	0.133	−0.008
	(0.026)	(0.845)	(0.008)
2014/15	0.047	0.480	0.001
	(0.036)	(0.700)	(0.006)
2016/17 (After)	0.017	0.789*	−0.010
	(0.023)	(0.370)	(0.012)
Coefficient comparisons (p-values in parenthesis)			
Q1 vs Q2	−0.018	−0.426	−0.016*
	(0.169)	(0.096)	(0.001)
Q1 vs Q3	−0.012	−0.384	−0.025**
	(0.424)	(0.243)	(0.000)
Q1 vs Q4	−0.015	−0.629	−0.033**
	(0.381)	(0.064)	(0.000)
Q1 vs Q5	−0.027	−1.098*	−0.033**
	(0.149)	(0.004)	(0.000)
Number of observations	511,343	511,343	511,343

Notes. ** $p<.01$, * $p<.0.05$. Models estimated using ordinary least squares and include hospital fixed effects and controls for Charlson co-morbidities, age in 10-year age bands and their interactions with sex. The full set of results displayed in the Online Supplementary Appendix Part E. For regression estimates, cluster robust standard errors in parentheses. For coefficient comparisons, p-values in parentheses. Q1 denotes the most income-deprived group, and Q5 the least income-deprived group.

3. Results

3.1. Descriptive statistics

Descriptive statistics are shown in Table 1. For the sample on Adult Critical Care in Trauma (TR1), 74 % of the 511,343 patients in the full sample were admitted to a hospital in the treatment group. The average discharge delay and the proportion of patients waiting more than four hours to be discharged from critical care were similar across treatment and control hospitals. The average discharge delay was 7.9 h. About half of patients (48 %) had a delay longer than 4 h to be discharged. The probability of a night discharge was also similar and experienced by 18 % of patients across the whole sample.

For the sample requiring an urgent CABG (IM1), 67 % of a total of 11,653 patients were admitted to hospital in the treatment group. The average in-hospital waiting time for CABG was slightly longer in the treatment group (10 days) than in the control group (9.4 days) where 52 % of patients waited >7 days for CABG, compared to 59 % in the treatment group. Length of stay was similar across the two groups and equal to 19.3 days across the whole sample. The risk of death within 30 and 365 days after CABG was low and equal to 2.7 % and 5.5 %, respectively, across the whole sample. The risk of experiencing a patient safety incident was 13 %.

For both schemes, more patients are from the most income-deprived areas of England. For CABG (IM1), this is more pronounced among patients admitted to hospitals in the treatment group compared to those in the control group, while for Adult Critical Care in Trauma (TR1) the income-deprivation distribution of patients is more similar between treatment and control groups.

We use age, sex and the Charlson co-morbidity index to measure severity. For the Adult Critical Care in Trauma (TR1) scheme, across the whole sample, 57 % of patients are male, while for urgent CABG (IM1),

Table 3
Regression results. Reducing cardiac surgery non-elective in-hospital waiting time (IM1).

	Waiting for CABG > 7 days	Wait for CABG (days)	Length of stay (days)	Death within 30 days	Death within 365 days	Patient safety incident
IMD Income quintile # Treated # Post						
Q1 # Treated # After	0.027	0.653	1.358*	0.011	0.017	−0.001
	(0.031)	(0.391)	(0.611)	(0.012)	(0.022)	(0.038)
Q2 # Treated # After	0.017	0.359	1.282	0.006	−0.003	0.014
	(0.039)	(0.422)	(0.635)	(0.012)	(0.018)	(0.036)
Q3 # Treated # After	0.027	0.215	0.676	−0.011	−0.031	−0.013
	(0.042)	(0.551)	(0.748)	(0.009)	(0.015)	(0.033)
Q4 # Treated # After	−0.045	−0.253	1.218	−0.007	−0.016	−0.000
	(0.042)	(0.518)	(0.609)	(0.008)	(0.016)	(0.032)
Q5 # Treated # After	0.036	0.090	1.337	−0.015	−0.023	−0.005
	(0.038)	(0.404)	(0.798)	(0.009)	(0.021)	(0.032)
Year # Treated						
2012/13 # Treated	0.115*	1.070	2.917**	−0.004	−0.006	0.041
	(0.053)	(0.631)	(1.028)	(0.010)	(0.012)	(0.029)
2013/14 # Treated	0.112	1.766	1.745	−0.009	−0.023	0.033
	(0.057)	(0.855)	(1.088)	(0.008)	(0.012)	(0.028)
2014/15 # Treated	0.079*	0.905	1.475*	0.007	−0.012	0.044*
	(0.033)	(0.505)	(0.644)	(0.008)	(0.014)	(0.017)
Year						
2012/13	0.035	0.710	−0.576	0.016	0.012	−0.014
	(0.043)	(0.502)	(0.746)	(0.009)	(0.010)	(0.022)
2013/14	0.011	0.090	0.008	0.013	0.025*	0.007
	(0.040)	(0.515)	(0.574)	(0.007)	(0.010)	(0.023)
2014/15	0.007	0.121	−0.237	0.001	0.015	−0.014
	(0.014)	(0.250)	(0.466)	(0.007)	(0.012)	(0.012)
2016/17 (After)	−0.044*	−0.812**	−2.118**	0.010*	0.012	0.018
	(0.017)	(0.221)	(0.435)	(0.005)	(0.011)	(0.030)
	(0.022)	(0.220)	(0.752)	(0.004)	(0.008)	(0.024)
Coefficient comparisons (p-values in parenthesis)						
Q1 vs Q2	−0.010	−0.294	−0.077	−0.005	−0.020	0.015
	(0.721)	(0.280)	(0.885)	(0.684)	(0.311)	(0.573)
Q1 vs Q3	−0.001	−0.438	−0.682	−0.022	−0.048	−0.012
	(0.993)	(0.480)	(0.342)	(0.094)	(0.042)	(0.625)
Q1 vs Q4	−0.073	−0.906	−0.140	−0.018	−0.033	0.001
	(0.179)	(0.086)	(0.790)	(0.137)	(0.072)	(0.972)
Q1 vs Q5	0.009	−0.563*	−0.021	−0.026*	−0.041	−0.004
	(0.821)	(0.031)	(0.977)	(0.032)	(0.104)	(0.867)
N. of observations	11,653	11,653	11,653	11,653	11,653	11,653

Notes. Models estimated using ordinary least squares and include hospital fixed effects and controls for Charlson co-morbidities, age in 10-year age bands and their interactions with sex. The full set of results displayed in the Online Supplementary Appendix Part E. For regression estimates, cluster robust standard errors in parentheses. For coefficient comparisons, p-values in parentheses. Q1 denotes the most income-deprived group.

79 % of patients are male. In both samples, most patients are 61 years or older. Across samples, half of all patients have two or more comorbidities. In the TR1 sample, 25 % have no comorbidities, while in the IM1 sample, only 10 % had no comorbidities.

3.2. Regression results

The estimation results from model (1) are shown in Tables 2 and 3. Table 2 is for Adult Critical Care in Trauma (TR1). It shows that the P4P incentive scheme for specialised care (PSS CQUIN) that aimed at reducing discharge delays in adult critical care was successful across all income quintiles. The effect is however heterogenous across socioeconomic groups. While patients from the most deprived areas experienced a reduction in their delay of discharge by 1.3 h (16.4 % at the sample mean) after the policy was introduced in the treatment group, the least deprived experienced a larger reduction in delay of discharge by 2.4 h (30.4 %). The effect shows a generally linear pattern across income-deprivation quintiles, except for quintiles 2 and 3 with similar reductions of 1.7 h (21.7 %). These results should be interpreted relative to an average increase in the post-policy period (2016/17) of discharge delay by 0.8 h across treatment and control groups.

The gradient can also be observed (albeit in a less linear way) for the policy target of reducing the probability of experiencing discharge delays of >4 h. Here, no statistically significant effect was found for patients from the most deprived areas, while the reduction was 8.6 percentage points for patients from the least deprived areas, and patients in quintiles 2–4 experienced reductions of between 7.1 and 7.7 percentage points. There was no statistically significant impact of the policy on night discharges across income-deprivation quintiles. Applying the Bonferroni correction for multiple comparisons, the estimates for Delayed discharge (hours) for Q2, Q4 and Q5 are still statistically significant, while the others are not. When applying instead the higher-powered sharpened False Discovery Rate (FDR) q-values [2], the Q3 result also remains statistically significant.

Comparing the estimates for Q1 to each of the other quintiles, we find that the reduction in discharge delay in hours was statistically significantly greater (in absolute values) for the least-deprived (Q5) than the most-deprived quintile (Q1), while the difference was not statistically significant when looking at the policy target of delays greater than 4 h. The reduction in the probability of a nightly discharge after the reform was statistically significantly higher (in absolute values) for all quintiles compared to most-deprived quintile.

Table 3 provides the results for urgent coronary bypass (IM1). It shows no statistically significant impact of the policy to reduce in-hospital waiting time for CABG on any of the outcomes studied across any of the income quintiles. The only exception is length of stay which increased for patients in the most deprived quintile at treatment group hospitals after the policy, but this result does not hold when the

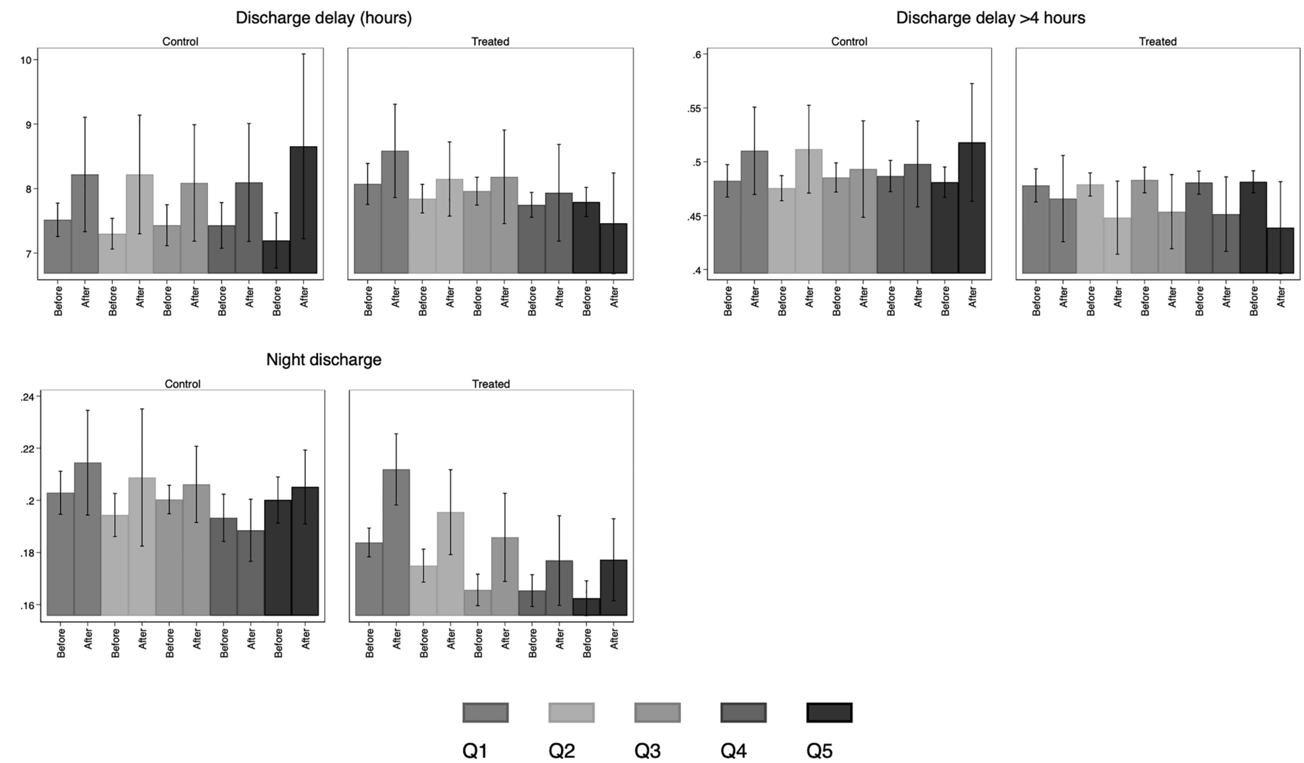

Fig. 1a. Adult critical care timely discharge (TR1).

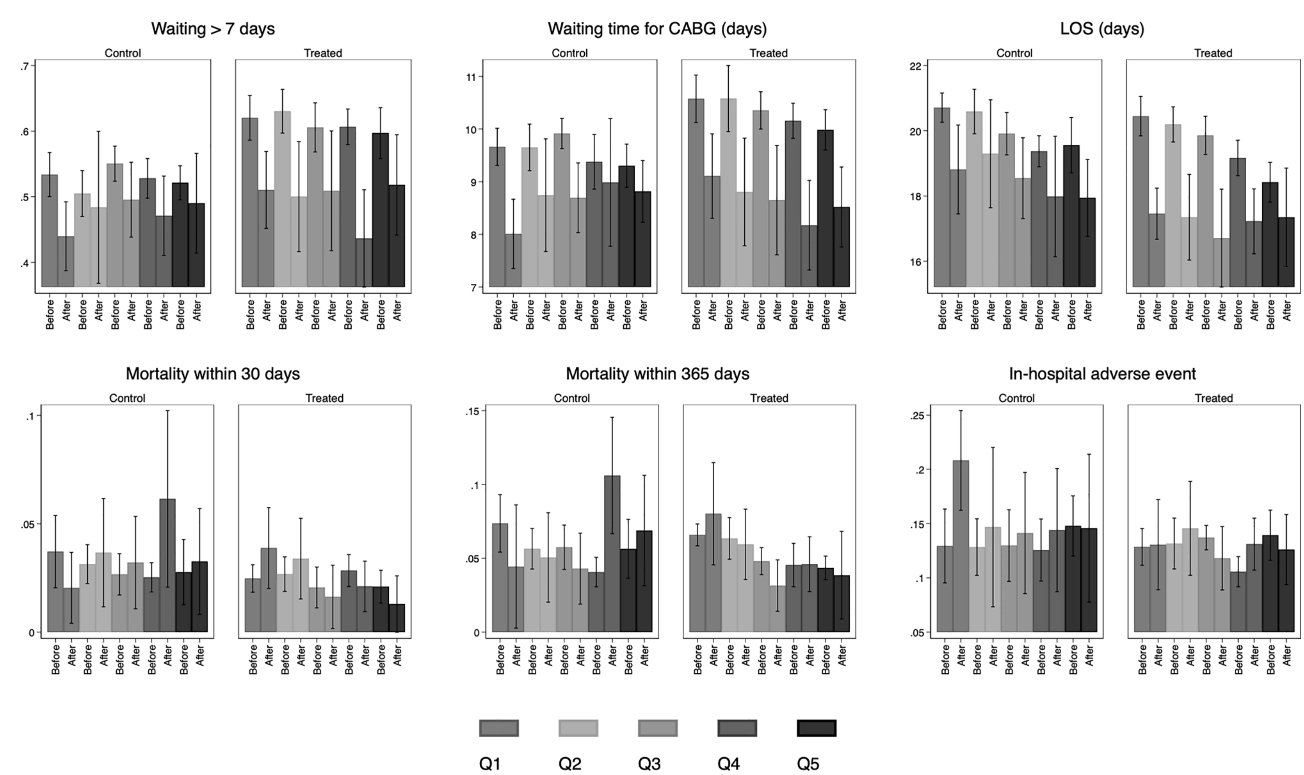

Fig. 1b. Reducing cardiac surgery non-elective in-hospital waiting time (IM1).

Bonferroni or FDR corrections are applied. Across both treated and control hospitals, there was an overall reduction in-hospital waiting time of 0.8 days in the year of the policy introduction compared to the year before, and the probability of waiting more than seven days for CABG fell by 4.4 percentage points while length of stay was two days lower overall. We find no statistically significant differences in the impact of the scheme across any quintiles compared to the most deprived (Q1) at 5 % significance level, except for mortality, where there was a greater reduction for Q3. Estimates from the event-study models are in line with the DiD estimates and displayed in the OSM

Part C. Parallel trend tests for Q1 and Q2-Q5 are shown in the OSM Part D. Full regression results for Table 3 are in OSM Part E.

Fig. 1 shows the predicted outcomes before and after scheme introduction and by hospital treatment status for each of the outcomes examined across income quintiles based on model (2). These figures give a more detailed representation of the socioeconomic gradient over time and across treatment groups. For example, it can be seen how the probability of experiencing a discharge delay from critical care of >4 h was similar across income quintiles and treated and control hospitals before the incentive scheme was introduced. After the scheme was introduced, the probability of experiencing a delay increased among all income quintiles in the control group, while it fell across all quintiles in the treatment group, with a larger decrease for the less deprived. Conversely, an already visible gradient in the probability of night discharge in treatment group hospitals in the before period was maintained when all patients experienced an increase in the night discharge probability in the after period. The figures also illustrate how the LOS reduction for CABG patients, which occurred across all groups from the before to after period, did seem to reduce inequalities in the treatment group, going from a visible socioeconomic gradient in LOS in the before period to a close to equal LOS in the after period Fig. 1b.

4. Discussion

Pay for performance is a policy lever that can be used to financially incentivise healthcare providers to improve quality. An immediate concern with this approach is that it may come at the cost of increasing healthcare inequalities. Our analysis suggests that for trauma care patients, P4P induced a larger reduction in delayed discharges amongst patients from less deprived areas, therefore widening inequalities. The difference across groups were in the order of magnitude of 1.7–2.4 h (16 %−30 % at the sample mean). However, we found no differences in the probability of being discharged during the night. Neither did we find any effect or differences in the effects of the scheme for patients requiring an urgent coronary bypass. Although the differences for trauma care are relatively small, the analysis suggests that the improvements arising from P4P may be larger for patients with higher socioeconomic status.

Our analysis also shows that there were limited differences in access to care by socioeconomic status before the introduction of the scheme. In the instances where we found differences, these were consistent with access to care being more favourable to less deprived patients, which is line with previous studies (e.g. [3,12,17,21]).

Our analysis has limitations. We focus on complex patients who require specialised care. Therefore, the results cannot necessarily be extrapolated to other settings. Given that the P4P scheme was voluntary, we have used propensity score matching to match treatment and control hospitals. We cannot rule out that residual selection effects remain.

There is scope to further refine the design of P4P schemes that include equity considerations. This could take the form of additional risk-adjustment to reduce variations in treatment costs by socioeconomic status [6] or adjusting tariffs to compensate for the additional cost of treating disadvantaged patients. More broadly, the assessment of P4P schemes requires to consider not only their cost-effectiveness [13] but the distributional implications [5].

5. Conclusion

There may be trade-offs between improving performance and inequalities of access to care. It is important that evaluations of P4P schemes assess more systematically the effect, not only on the average performance but also across different socioeconomic groups.

Funding

This research is funded by the National Institute for Health and Care Research (NIHR) Policy Research Programme (project reference PR-R18–0117–22,001) and the National Institute for Health Research (NIHR) Imperial Patient Safety Translational Research Centre (grant number PSTRC-2016–004). The views expressed are those of the authors and not necessarily those of the NIHR or the English Department of Health and Social Care.

CRediT authorship contribution statement

Alberto Núñez-Elvira: Data curation, Formal analysis, Methodology, Writing – review & editing. **Yan Feng:** Conceptualization, Methodology, Project administration, Writing – review & editing. **Søren Rud Kristensen:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft. **Paula Lorgelly:** Conceptualization, Writing – review & editing. **Rachel Meacock:** Conceptualization, Writing – review & editing. **Luigi Siciliani:** Conceptualization, Methodology, Writing – original draft. **Matt Sutton:** Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.healthpol.2025.105365.

References

[1] Alshamsan R, Majeed A, Ashworth M, Car J, Millett C. Impact of pay for performance on inequalities in health care: systematic review. J Health Serv Res Policy 2010;15(3):178–84.
[2] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B Methodol 1995;57:289–300.
[3] Bosque-Mercader L, Carrilero N, García-Altés A, López-Casasnovas G, Siciliani L. Socioeconomic inequalities in waiting times for planned and cancer surgery: evidence from Spain. Health Econ 2023;32(5):1181–201.
[4] Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373–83.
[5] Cookson R, Griffin S, Norheim OF, Culyer AJ. Distributional cost-effectiveness analysis: quantifying health equity impacts and trade-offs. Oxford University Press; 2020.
[6] Conway A, Satin D. The role of pay-for-performance in reducing healthcare disparities: a narrative literature review. Prev Med 2022;164:107274.
[7] Feng Y, Kristensen SR, Lorgelly P, Meacock R, Sanchez MR, Siciliani L, Sutton M. Pay for performance for specialised care in England: strengths and weaknesses. Health Policy 2019;123(11):1036–41 (New York).
[8] Feng Y, Kristensen SR, Lorgelly P, Meacock R, Núñez-Elvira A, Rodés-Sánchez M, Siciliani L, Sutton M. Pay-for-performance incentives for specialised services in England: a mixed methods evaluation. Eur J Health Econ 2024;25(5):857–76.
[9] Kaarboe O, Siciliani L. Contracts for primary and secondary care physicians and equity-efficiency trade-offs. J Health Econ 2023;87:102715.
[10] Marmot M. Fair society, healthy lives: the Marmot review; strategic review of health inequalities in England post-2010. Marmot Rev 2010;2010.
[11] Mason T, Lau YS, Sutton M. Is the distribution of care quality provided under pay-for-performance equitable? Evidence from the Advancing Quality programme in England. Int J Equity Health 2016;15:1–6.
[12] Matias MA, Santos R, Siciliani L, Sivey P, Proctor A. Socioeconomic inequalities in waiting times for breast cancer surgery. Health Econ 2024. forthcoming.
[13] Meacock R, Kristensen SR, Sutton M. The cost-effectiveness of using financial incentives to improve provider quality: a framework and application. Health Econ 2014;23(1):1–13.
[14] Mendelson A, Kondo K, Damberg C, Low A, Motúapuaka M, Freeman M, O'Neil M, Relevo R, Kansagara D. The effects of pay-for-performance programs on health, health care use, and processes of care: a systematic review. Ann Intern Med 2017;166(5):341–53.
[15] Milstein R, Schreyoegg J. Pay for performance in the inpatient sector: a review of 34 P4P programs in 14 OECD countries. Health Policy 2016;120(10):1125–40 (New York).
[16] Miraldo M, Siciliani L, Street A. Price adjustment in the hospital sector. J Health Econ 2011;30(1):112–25.
[17] Moscelli G, Siciliani L, Gutacker N, Cookson R. Socioeconomic inequality of access to healthcare: does choice explain the gradient? J Health Econ 2018;57:290–314.
[18] England N.H.S., 2016. IM1Reducing Cardiac surgery non-elective inpatient waiting https://www.england.nhs.uk/wp-content/uploads/2016/03/im1-reduc-cardiac-surgry-non.pdf.

[19] Parkinson B, McManus E, Sutton M, Meacock R. Does recruiting patients to diabetes prevention programmes via primary care reinforce existing inequalities in care provision between general practices? A retrospective observational study. BMJ Qual Saf 2023;32(5):274–85.
[20] Siciliani L, Straume OR. Competition and equity in health Care markets. J Health Econ 2019;64:1–14.
[21] Simonsen NF, Oxholm AS, Kristensen SR, Siciliani L. What explains differences in waiting times for health care across socioeconomic status? Health Econ 2020;29 (12):1764–85.
[22] Turner AJ, Francetic I, Watkinson R, Gillibrand S, Sutton M. Socioeconomic inequality in access to timely and appropriate care in emergency departments. J Health Econ 2022;85:102668.
[23] Wright A, Cornick E, Jones H, Marshall L, Roberts A. Towards an effective NHS payment system: eight principles. London: The Health Foundation; 2017.
[24] Zaresani A, Scott A. Is the evidence on the effectiveness of pay for performance schemes in healthcare changing? Evidence from a meta-regression analysis. BMC Health Serv Res 2021;21:1–10.