
Article:
Gaughan, James Michael orcid.org/0000-0002-8409-140X, Gutacker, Nils orcid.org/0000-0002-2833-0621, Grasic, Katja et al. (3 more authors) (2019) Paying for Efficiency:Incentivising same-day discharge in the English NHS. Journal of health economics. ISSN: 0167-6296

https://doi.org/10.1016/j.jhealeco.2019.102226

Paying for Efficiency: Incentivising Same-Day Discharges in the English NHS

James Gaughan*[1], Nils Gutacker[1], Katja Grašič[1], Noemi Kreif[1], Luigi Siciliani[2], and Andrew Street[3]

[1]Centre for Health Economics, University of York
[2]Department of Economics and Related Studies, University of York
[3]Department of Health Policy, The London School of Economics and Political Science

Abstract

We study a pay-for-efficiency scheme that encourages hospitals to admit and discharge patients on the same calendar day when clinically appropriate. Since 2010, hospitals in the English NHS are incentivised by a higher price for patients treated as same-day discharge than for overnight stays, despite the former being less costly. We analyse administrative data for patients treated during 2006-2014 for 191 conditions for which same-day discharge is clinically appropriate – of which 32 are incentivised. Using difference-in-difference and synthetic control methods, we find that the policy had generally a positive impact with a statistically significant effect in 14 out of the 32 conditions. The median elasticity is 0.24 for planned and 0.01 for emergency conditions. Condition-specific design features explain some, but not all, of the differential responses.

JEL: D22, I11

Keywords: Pay for Performance; prospective payment systems; activity based funding; hospital incentives; DRGs; synthetic control method; policy evaluation

*Corresponding author. Centre for Health Economics, University of York, Heslington, UK; tel: +44 1904 321423; E-Mail: james.gaughan@york.ac.uk

1 Introduction

Many healthcare systems reimburse hospitals through prospective payment systems (PPS) in which the price for a defined unit of activity, such as a Diagnosis Related Group (DRG), is set in advance and is equal across hospitals (Paris et al. 2010). Economic theory predicts that hospitals will expand activity in areas where price exceeds marginal costs and minimise activity in areas where they stand to make a loss.[1] This form of reimbursement should encourage hospitals to engage in efficient care processes and cost reduction strategies to improve profit margins (Shleifer 1985; Ellis and McGuire 1986; Ma 1994; Hodgkin and McGuire 1994).

One way to reduce costs is by reducing length of stay (LoS), this being an important cost driver. For some patients it may be possible to reduce overnight stays to zero, specifically those for whom care can be provided safely[2] within a setting in which patients are admitted, treated and discharged on the same day ('*same day discharge*' (SDD)). Not only may an SDD be less costly to provide, it might also be beneficial to some patients if they can recover in the comfort of their own home and are less exposed to potentially infectious hospital environments. Increasing SDDs for these patients generates a welfare improvement driven by lower provider costs and unaltered or improved health benefits for patients. The British Association of Day Surgery (BADS) (2006) has recommended the adoption of SDD for 157 types of planned surgery and the British Association for Ambulatory Emergency Care (BAAEC) (2014) has identified a range of 34 conditions that require urgent care but where a subsequent overnight stay for observation is generally considered unnecessary. Implementing these recommendations is also in the financial interest of hospitals reimbursed according to the English form of PPS, which pays the same amount for SDD admissions and for admissions with an overnight hospital stay, despite the cost of providing SDD care being lower (Street and Maynard 2007).[3] Therefore, hospitals can improve profits by increasing the proportion of patients treated on an SDD basis rather than keeping them in hospital overnight.

Despite these recommendations and financial incentives, SDD rates are lower than is clinically recommended for a wide range of conditions (Department of Health 2009)(see also Figure 1). The

[1] (Semi-)altruistic providers may be willing to treat patients for which marginal costs exceed price as long as the financial losses are offset by sufficient patient benefit. The extent to which this is possible depends on the potential for cross-subsidisation within the organisation, and whether they face a soft budget constraint (Brekke et al. 2015).

[2] As early as 1985, the Royal College of Surgeons of England (1985) noted that "*it should be clear to all concerned, the surgeon, the nursing staff, and in particular the patient, that day-surgery is in no way inferior to conventional admission for those procedures for which it is appropriate, indeed it is better.*" (Royal College of Surgeons of England 1985).

[3] For example, in 2013/14 the average cost of planned surgery carried out as a day case in the English NHS was £698 compared to the average cost of £3,375 for overnight stays. (`https://www.kingsfund.org.uk/blog/2015/07/day-case-surgery-good-news-story-nhs`)

reasons for these low rates may relate to financial constraints on hospitals that limit their ability to invest in dedicated same-day facilities or reluctance by doctors to change established working practices. One way to encourage hospitals and doctors to increase uptake of SDD care is to increase the SDD price. This has been the approach taken in England under a payment reform known as the SDD bonus policy (Monitor & NHS England 2014). Hospitals receive an SDD bonus on top of the base DRG price for treating a patient as an SDD compared to an overnight admission. Starting in 2010, the reform has been progressively applied to 32 different conditions.

Our analysis of this policy reform makes two main contributions to the literature. First, it contributes to our understanding of economic incentives in the health sector by exploiting unique features of the SDD policy that relate to the economic importance of the bonus and the focus on efficiency (as opposed to other dimensions such as quality or overall volume). It is designed to incentivise technical efficiency, by paying hospitals extra to reduce length of stay and use of care inputs, such as staff time and hospital beds, by shifting care delivery from more expensive overnight wards to less costly same day settings. A distinctive feature of the SDD bonus policy is that the incentive scheme is high-powered, in that it pays more for the less costly SDD treatment. This contrasts with the common form of PPS in which prices are set at average cost (Shleifer 1985), either pooled across SDD and overnight stay (e.g. as in England), or separately for each admission type (e.g. as in Norway where the price is lower for an SDD than an overnight stay in line with the different average costs). In England, the cost advantage varies across the 32 conditions from 23% to 71% lower for SDD than for an overnight hospital stay in the pre-policy period. The SDD bonus compounds this advantage and is also economically significant, varying from 8% to 66% more than for an overnight stay. We are able to exploit this heterogeneity in the size of the incentive to assess whether it predicts changes in behaviour.

We also contribute to analytical studies that employ relatively new synthetic control (SC) methods and compare these to more traditional difference-in-difference (DID) methods. To evaluate the effectiveness of the policy we exploit the fact that incentives have been applied to 32 conditions, using non-incentivised conditions as control groups. SC methods are a potentially useful addition to the analytical armoury in situations where it is possible to draw on a large number of potential control groups. Following the pioneering work by Abadie and Gardeazabal (2003) and Abadie et al. (2010), SC methods are receiving increasing attention in the wider economic literature (Billmeier and Nannicini 2013; Bharadwaj et al. 2014; Green et al. 2014; Acemoglu et al. 2017). Within health economics, SC methods have been applied to study the effect of co-payments (Olsen and Melberg

2018), tax incentives (Fletcher et al. 2015; Bilgel and Galle 2015), public health interventions such as malaria eradication (Barofsky et al. 2015), and expansion of health insurance (Hu et al. 2018; Hernæs 2018). SC methods have been very rarely applied to provider incentives. We are only aware of one study by Kreif et al. (2016), which applies SC methods to evaluate the effect of a regional pay-for-performance (P4P) scheme in England on mortality rates. These studies all consider a single policy initiative with associated idiosyncrasies, which provides limited evidence on the general applicability of SC methods for policy evaluations typically considered in health economics. In contrast, we evaluate 32 policy variants of a particular payment reform following a common analysis plan (e.g. sample period, unit of assessment, criteria for selecting suitable control groups, etc.). This yields insights into whether DID and SC methods generate consistent conclusions in terms of point estimates and statistical inference under a range of different scenarios.

Our key findings on the effectiveness of the policy are as follows. We find that the policy led to a statistically significant increase in SDD rates of 5 percentage points (pp) for planned conditions and 1pp for emergency conditions. However, there is considerable heterogeneity across conditions with eight out of 13 planned conditions showing statistically significant positive effects in DID analysis. Estimated effects range from -2 to +22pp changes in SDD rates. Results are more mixed for emergency conditions, where we find that the policy had a statistically significant positive effect on six out of 19 emergency conditions but caused reductions in SDD rates for two conditions. The range of estimated effects is also narrower (-6 to +6pp) and more centred around zero. The median elasticity of SDD rates to price is 0.24 for planned conditions and 0.01 for emergency conditions (overall median = 0.09). Elasticities are larger for conditions with larger post-policy price differences between SDD and overnight care, and, for planned conditions only, with bigger profit margins. In relation to the methods employed, our analysis suggests that DID and SC methods provide similar point estimates when there is a large pool of potential control conditions to choose from, as is the case for planned conditions. However, even in such favourable instances, inference from SC methods are still considerably more conservative, resulting in fewer statistically significant findings than in DID analysis.

Our analysis relates to two strands of the literature within the broader area of hospital incentive schemes (Chandra et al. 2011). First, we contribute to studies that focus on the effect of changes in prices designed to encourage hospitals to reduce LoS. It is well established that PPS encourages reductions in LoS compared to either fee-for-service or global budgeting arrangements, by making hospitals more cost-conscious than the alternative funding regimes. This was examined in pioneering

work by Rosko and Broyles (1986), Salkever et al. (1986), Long et al. (1987), and Lave and Frank (1990) and others in the US Medicare and Medicaid systems, and has subsequently been confirmed in a range of other countries (e.g. Shmueli et al. 2002; Farrar et al. 2009; Moreno-Serra and Wagstaff 2010; O'Reilly et al. 2012). As well as finding general reductions in LoS, Farrar et al. (2009) estimated that the introduction of PPS in the English NHS led to an 0.4 to 0.8% increase in SDD rates for planned surgery. Much less is known about the ability of payers to influence LoS through deliberate price setting within a PPS arrangement. Shin (2019) exploits the 2005 Medicare change in its definition of payment areas that generated exogenous area-specific price shocks. The study found that the higher price did not affect volume, LoS and quality of services but it induced shifting patients into higher-paying DRGs. This is in line with Dafny (2005), who found that a 10% increase in price due to the removal of an age criterion in the allocation of patients to DRGs led to upcoding without significant change in LoS. Verzulli et al. (2017) study the effect of a one-time price increase for a subset of DRGs in the Emilia-Romagna region of Italy. They find evidence that hospitals expand the provision of surgery in response to more generous reimbursement but this has no effect on waiting times or LoS. More closely related to our setting, Januleviciute et al. (2016) examine the choice of SDD care versus overnight stay in the Norwegian context, where prices are differentiated by admission type. They find no evidence that hospitals respond to intertemporal variation in the price mark-ups for overnight stays relative to SDD care by changing their discharge practice.

In none of the above-mentioned settings were prices set with the explicit aim to reduce LoS. A noteworthy exception is the study by Allen et al. (2016), who considered the impact of the SDD bonus policy in England on a single incentivised condition, cholecystectomy, within a DID framework with a control group of all non-incentivised procedures recommended for SDD care. This study found an increase in SDD rates of 5.8 percentage points in the first 12 months following the policy introduction. As well as comparing DID and SC methods, we extend this earlier analysis to 31 additional conditions, allowing us to examine the generalisability of the previous result and study the determinants of the potentially heterogeneous responses to the SDD bonus. Furthermore, we examine longer-term effects, up to five years after the introduction of the bonus, allowing us to examine whether short-term effects are maintained over time.

Our study also contributes to a second strand of literature evaluating P4P programmes. A recent study reviews 34 hospital sector P4P schemes in high-income countries (Milstein and Schreyögg 2016). Most of the P4P schemes reviewed focus on incentivising quality, either through rewarding health outcomes or process measures of quality, and involve small or moderate bonuses of 5% or less

(Cashin et al. 2014). Effects are generally modest in size, short-lived and sometimes associated with unintended consequences. In contrast to the existing P4P literature, the policy we evaluate has two distinct features. First, few P4P schemes incentivise technical efficiency directly, so this study contributes to the small literature on what we label "pay-for-efficiency" (P4E) schemes. Second, the SDD bonus policy is much more high-powered than previous P4P schemes and, therefore, our analysis can shed light on whether limited responsiveness to P4P schemes as documented in the literature is simply due to insufficient financial incentive, as has been hypothesised (Milstein and Schreyögg 2016).

The study is organised as follows. Section 2 provides the institutional background and the SDD pricing policy. Section 3 describes the data. Section 4 outlines the empirical methods. Section 5 describes the results. Section 6 is devoted to discussion and concluding remarks.

2 Institutional background and behavioural predictions

The English NHS is funded by general taxation and residents have to be registered with a general practitioner. There are two routes to hospital: either patients are referred by their general practitioner for care 'planned' in advance (e.g. scheduled surgery) or they are admitted for immediate 'emergency' care after attending the hospital's emergency department. The SDD bonus policy applies to both planned and emergency conditions. NHS patients face no charges for hospital care, whether in publicly owned NHS hospitals or the small number of private hospitals that provide care to NHS patients. All NHS hospital doctors are salaried and do not share in hospitals' profits or losses.

The NHS adopted a PPS for hospital reimbursement in 2003. Hospitals are paid a pre-determined price for treating NHS-funded patients, differentiated by Healthcare Resource Groups (HRGs; the English equivalent of DRGs). Patients are assigned to a HRG based on diagnoses, procedures and, in some cases, other characteristics such as age (Department of Health 2002; Grašič et al. 2015). Initially limited to a small number of planned conditions, PPS has been extended progressively over time and now covers most hospital activity.

Before the SDD policy was introduced, the HRG payment was the same for both same day and for overnight stays across planned treatments[4]. This was not the case for emergency care, where

[4] Hospitals also receive additional per diem payments for each additional night a patient stays in hospital beyond a HRG-specific long-stay trim point. This trim point is set at the 75th percentile plus 1.5x the interquartile range of the LoS distribution in the HRG. Such long-stay adjustments are not relevant to our study since the SDD policy is directed at the low end of the LoS distribution.

the payment for same day treatments was lower than for overnight stays (to reduce the incentive to admit less severe patients for overnight observation).

From 2010, the English Department of Health has gradually introduced explicit incentives in the form of the SDD bonuses, which give a stronger financial incentive to reduce LoS. For patients allocated to the same HRG, the policy involved increasing the payment for someone treated on an SDD basis, with an offsetting reduction in the base HRG price for those who stay overnight. The difference between these two prices constitutes the SDD bonus. The specific conditions to which the SDD bonuses apply are drawn from a list compiled by the British Associations of Day Surgery and for Ambulatory Emergency Care for which overnight stay is considered unnecessary and where there is clinical consensus about the appropriate level of SDD.[5] The BADS and BAAEC both produce directories listing 191 clinical conditions (i.e. specific diagnoses or surgical treatments) between them that are deemed suitable for SDD with recommended rates (RRs) of SDD that are considered safe and appropriate (British Association of Day Surgery 2006; British Association for Ambulatory Emergency Care 2014).

The SDD bonuses apply to all public and private hospitals providing publicly-funded care. The selection and design of the bonuses was informed by discussions with clinical stakeholders and varies across clinical areas (Department of Health 2007). The general criteria for potential selection are volume (>5,000 patients/year)[6], the national SDD rate being below the RR for this condition, and evidence of variation in the SDD rate across hospitals (Department of Health 2009). Not all clinical conditions meeting these general criteria have an SDD bonus but by April 2014, 13 planned and 19 emergency conditions were covered by the incentive scheme (Monitor & NHS England 2014). To qualify for the bonus payment, the patient has to be admitted and discharged on the same day. In addition, for planned treatments, the care has to be scheduled as SDD in advance of admission. New conditions to be incentivised are announced six months in advance of introduction.

Since the introduction of the SDD bonus policy the price for same day discharge is systematically higher than for overnight stay across the 32 SDD conditions. As an example, in 2010 hospitals were paid £329 (or 24%) more for cholecystectomy (gall bladder removal) provided as SDD (Department of Health 2009). The absolute and relative size of the price differential varies considerably across

[5] In some cases, additional exclusion criteria are applied to limit the scope of the SDD bonus to non-complex patients. In these cases, the group of patients with incentivised prices attached is a subset of those given in relevant directories and recommended rates can be considered a lower bound of what is clinically appropriate.

[6] An exception is 'simple mastectomy' which has been incentivised since 2011 despite an annual volume of about 4,000 patients.

the 32 incentivised conditions, ranging from 8% to 66% of the overnight admission price. Once introduced, bonus differentials are fairly stable over time[7].

Table 1 provides an overview of the incentivised SDD conditions, the financial year in which the incentive was introduced[8], the price with and without the SDD incentive, the average cost of care reported by NHS hospitals in the year prior to the policy, as well as the SDD rate and the number of patients eligible in the twelve months prior to announcement of the incentive for that condition.

Notice that in the pre-policy period hospitals already had a financial incentive to treat planned patients as SDD up to the recommended rate given that the cost of SDD is nearly always lower than the cost of an overnight stay. But as shown below in Section 3, hospitals had very low planned SDD rates in the pre-policy period, and always well below the RR. This could be due to the motivations of the doctor providing treatment or the constraining features of the hospital in which the doctor works, which we discuss in turn.

As regards low motivation, slow uptake of SDD may reflect poor dissemination about best practice. Doctors may have established practices and be reluctant to engage in disruptive innovations or simply may not be aware of or doubt the evidence that SDD is as safe as traditional practice involving overnight admission for the conditions concerned. They may also struggle to identify the patient population that is suitable for SDD, particularly if it is not recommended for all patients, i.e. RR $< 100\%$. Greater uptake of SDD may also require some re-training (e.g. in laparoscopic surgical techniques) that carries monetary and time costs for doctors.

The hospital in which the doctor works may be constrained in its ability to extend SDD to more patients. To a limited extent, SDD treatments can be offered in a normal hospital setting. However, scaling-up the provision of SDD treatment requires dedicated physical space and facilities. The hospital may have to invest in a dedicated facility, either by opening up new buildings or by engaging in re-organisation of existing wards. This would involve fixed costs which would be justifiable to senior managers only if it offers the prospect of long-term financial returns. Hospitals may not undertake this investment, particularly if they face borrowing constraints that restrict their access to capital funds (Marini et al. 2008; Thompson and McKee 2011). Moreover, managers faced with the various day-to-day issues of running a hospital may find it difficult to allocate the necessary time and resources to engage in more strategic re-organisations. Paying a bonus for

[7] The bonus as a percentage of base price changed by more than 5% from introduction to the financial year 2014/15 for six out of 32 SDD conditions. This variation arises due to changes to the base price that reflects year-on-year variation in the reported cost data used for price setting rather than because of purposeful policy refinement.

[8] Financial years run from 1st April to 31st March of the following calendar year.

Table 1: Overview of incentivised conditions

#	Condition	Year of introduction	Recommended rate (RR)(%)	Number of patients eligible (pre-policy)	SDD rate (pre-policy) (%)	Price						Production cost (pre-policy)		
						Pre-policy			Post-policy					
						SDD	ON	Δ	SDD	ON	Δ	SDD	ON	Δ
Planned care														
1	Cholecystectomy	2010	60	11,004	16	1,365	1,365	0	1,694	1,369	325	1,365	2,145	-780
2	Simple mastectomy	2011	15	4,048	7	2,123	2,123	0	2,385	2,085	300	1,480	2,682	-1,202
3	Sentinel node mapping and resection	2011	80	13,971	31	2,073	2,073	0	1,376	1,076	300	1,423	2,574	-1,151
4	Operations to manage female incontinence	2011	80	13,658	25	1,222	1,222	0	995	695	300	1,021	1,574	-553
5	Endoscopic prostate resection	2011	15	6,395	1	1,959	1,959	0	1,947	1,797	150	1,274	2,321	-1,047
6	Laser prostate resection	2011	90	16,000	3	1,890	1,890	0	1,863	1,563	300	1,240	2,236	-996
7	Hernia repair	2011	85	90,575	57	1,233	1,233	0	1,124	824	300	1,287	1,913	-626
8	Therapeutic arthroscopy of shoulder	2011	80	26,836	49	2,172	2,172	0	2,253	2,053	200	1,319	2,047	-729
9	Bunion operations	2011	85	16,148	50	1,063	1,063	0	1,170	970	200	1,123	1,972	-848
10	Fasciectomy	2011	95	9,211	74	2,735	2,735	0	2,297	2,097	200	1,499	2,286	-787
11	Tonsillectomy	2012	80	15,243	37	1,074	1,074	0	1,071	771	300	1,130	1,468	-337
12	Septoplasty	2012	80	18,830	48	1,164	1,164	0	1,204	1,004	200	1,219	1,622	-403
13	Tympanoplasty	2013	80	7,577	48	2,008	2,008	0	2,182	1,882	300	2,038	2,947	-909
Emergency care														
14	Epileptic seizure	2012	90	42,601	27	445	1,781	-1,336	1,157	946	211	435	1,713	-1,278
15	Acute headache	2012	60	55,826	34	511	730	-219	748	537	211	424	1,151	-727
16	Asthma	2012	30	27,986	23	606	1,173	-568	1,081	891	190	404	1,190	-785
17	Lower respiratory tract infections without COPD	2012	60	9,794	40	489	1,086	-597	776	585	191	412	1,137	-725
18	Pulmonary embolism	2012	90	11,235	14	512	2,049	-1,536	1,658	1,468	190	476	1,697	-1,221
19	Chest pain	2012	60	232,317	41	561	802	-241	748	543	205	433	1,216	-783
20	Appendicular fractures not requiring fixation	2012	60	39,931	30	298	1,111	-813	832	599	233	554	2,262	-1,708
21	Cellulitis	2012	90	28,965	25	568	1,477	-909	1,147	924	222	433	1,546	-1,113
22	Renal/ureteric stones	2012	60	28,241	33	642	876	-234	821	606	215	459	1,273	-814
23	Deep vein thrombosis	2012	90	18,121	56	612	1,360	-748	785	558	227	463	1,718	-1,255
24	Deliberate self-harm	2012	90	95,973	46	414	532	-119	535	326	209	372	899	-527
25	Falls including syncope or collapse	2012	90	62,230	32	443	985	-542	751	546	205	401	994	-593
26	Community acquired pneumonia	2013	30	11,121	19	609	1,353	-744	1,136	936	200	447	1,374	-927
27	Arrhythmia	2013	60	96,203	26	682	1,588	-906	1,242	1,026	216	465	1,373	-908
28	Minor head injury	2013	60	13,976	53	477	546	-69	698	453	245	424	1,074	-649
29	Low risk pubic rami	2013	90	6,935	8	344	1,374	-1,030	1,711	1,466	245	971	3,861	-2,890
30	Bladder outflow obstruction	2013	60	11,133	23	632	1,121	-489	1,009	798	211	423	1,373	-950
31	Anaemia	2013	90	13,315	16	635	2,249	-1,614	1,908	1,662	246	525	1,440	-915
32	Abdominal pain	2013	60	199,320	31	441	441	0	918	693	225	452	452	0

SDD = Same day discharge; ON = Overnight

Note: If incentive applied to more than one HRG within a condition, the price and cost information shown are weighted averages according to volume.

Pre- and post-policy refer to the 12 months before or after the policy start, respectively. The pre-policy SDD rate is calculated in the 12 months prior to the policy announcement and therefore not affected by anticipatory effects.

activity conducted on an SDD basis may be sufficient to overcome both clinical and managerial resistance.

More formally, denote the pre-policy period with $\alpha = 0$ and the post-policy period as $\alpha = 1$. The price for a HRG (g) in year (k) in the pre-policy period ($P_{0,k,g}$) is proportional to the average cost of care reported across all English NHS hospitals for patients (admitted as planned or emergency) who were treated three years before, $\bar{C}_{k-3,g} = \sum_{j=1}^{J}(C_{k-3,j,g} \times N_{k-3,j,g}) / \sum_{j=1}^{J} N_{k-3,j,g}$, where $j = 1 \ldots J$ denotes the hospital, $N_{k-3,j,g}$ is the number of patients for a given hospital j, and $C_{k-3,j,g}$ is the average cost of patients in hospital j[9]. Prices are further adjusted to account for inflation (I) and expected general technical efficiency improvement (E) factors[10] . Therefore, the pre-policy price is $P_{0,k,g} = \bar{C}_{k-3,g} \times I_k \times E_k$ with $I_k > 1$ and $E_k < 1$. For most planned treatments, hospitals are paid the same for patients admitted and discharged on the same day (SDD) or overnight stays (ON). Therefore, $P_{0,k,g} = P^{SDD}_{0,k,g} = P^{ON}_{0,k,g}$ if treatment is planned. However, a short-stay adjustment is applied to patients admitted as an emergency and discharged on the same day. The adjustment takes the form of a factor $0 \leq \lambda \leq 1$ which takes the value 1 if the national average length of stay for the HRG is less or equal to two nights and increasingly smaller values as average length of stay increases. Therefore, emergency care including at least one overnight stay has a price constructed equivalently to planned care $P^{ON}_{0,k,g} = P_{0,k,g}$ while $P^{SDD}_{0,k,g} = \lambda P_{0,k,g}$.

We compare the financial incentives that hospitals faced before and after the policy. To keep the presentation simple, we suppress the HRG and year notation (g and k) and also assume that (i) each hospital has a total volume of patients treated (either as SDD or overnight) equal to N and that this is constant over time, (ii) each hospital has identical costs, therefore also suppressing j, but average costs can vary over time before and after the policy (for example as a result of the change in case-mix arising from a change in the proportion of patients treated as overnight admission).

In summary, the price pre-policy is P_0 and post-policy is P^{SDD}_1 for same-day discharge and P^{ON}_1 for an overnight stay. Hospital incentives are driven not only by differences in prices but also differences in costs. Define C^{ON}_0 and C^{SDD}_0 as respectively the average cost of an overnight stay and a same-day discharge in the pre-policy period (and C^{ON}_1, C^{SDD}_1 in the post-policy period).

[9] All NHS hospitals provide detailed reference cost information to the Department of Health on an annual basis. These data are collated in the reference cost schedule and provide information on the average cost of production across hospitals, further broken down by admission type.

[10] The base price is further adjusted for hospital-specific factors such as local cost of capital and labour and specialist hospital status. As the policy evaluated is national and applies equally to all hospitals, these hospital-specific adjustments do not affect the incentives created.

The profit function for planned SDD activity, denoted π, in the pre-policy and the post-policy period is given respectively by

$$\pi_0 = N_0^{SDD}(P_0 - C_0^{SDD}) + (N - N_0^{SDD})(P_0 - C_0^{ON}) \tag{1}$$

$$\pi_1 = N_1^{SDD}(P_1^{SDD} - C_1^{SDD}) + (N - N_1^{SDD})(P_1^{ON} - C_1^{ON}) \tag{2}$$

and the difference in profit before and after the policy is:

$$\begin{aligned}\Delta\pi = \pi_1 - \pi_0 = (P_1^{SDD} - P_1^{ON})N_1^{SDD} - N(P_0 - P_1^{ON}) \\ + (N_1^{SDD} - N_0^{SDD})(C_0^{ON} - C_0^{SDD}) \\ - [N_1^{SDD}(C_1^{SDD} - C_0^{SDD}) + (N - N_1^{SDD})(C_1^{ON} - C_0^{ON})]\end{aligned} \tag{3}$$

Under the assumptions outlined above, the first term is positive and gives the additional revenues for every treatment which is provided as SDD. The second term is negative and is given by the reduction in revenues due to a reduction in the overnight price. The third term is positive if the SDD price induces an increase in the SDD rate, which is less costly (evaluated at pre-policy costs). The fourth and last term, in square brackets, relates to changes in the average costs, which can be due to patient composition or external factors, the sign being generally indeterminate. We could argue, for example, that patients who are treated as SDD after the policy are at the margin more severe, so that this will translate into an increase in the average cost of SDD and a reduction in the average cost of an overnight stay (see Siciliani (2006) and Hafsteinsdottir and Siciliani (2010) for more formal theoretical models). However, we assume that the increase in average costs for SDD is relatively small, so that an increase in SDD rates leads to a reduction in overall costs (i.e. the sum of the third and fourth term is positive).

The analysis highlights that the SDD pricing policy generates a financial incentive for hospitals, equal to $P_1^{SDD} - P_1^{ON} > P_0^{SDD} - P_0^{ON} > 0$, to increase planned SDD treatments, but the overall effect on profits also depends on the reduction in the base price. A similar analysis holds for emergency care where the only difference is that pre-policy the price was higher for overnight treatments, i.e. $P_1^{SDD} - P_1^{ON} > P_0^{SDD} - P_0^{ON} < 0$.

Differentiating equation 3 with respect to the number of SDD treatments, N_1^{SDD}, we obtain the financial incentive to treat an additional patient as an SDD. This is given by $(P_1^{SDD} - P_1^{ON}) - (C_1^{SDD} - C_1^{ON})$, which is always positive whenever the cost of SDD activity is lower than the cost of

an overnight admission. The expression suggests that, potentially, hospitals have a strong financial incentive to increase the number of SDD patients.

3 Data

We use data from Hospital Episode Statistics (HES) on all NHS-funded patients aged 19 or older admitted to English hospitals between April 2006 and March 2015 for care which could be delivered as SDD according to the BADS / BAAEC directories (157 planned and 34 emergency conditions). HES is an admission-level dataset that contains detailed information on patients' clinical and socio-demographic characteristics, the admission pathway and its timings, and whether care was scheduled as SDD in advance (planned admissions only). A patient is considered to have received SDD care if admission and discharge date coincide.

Figure 1 shows the SDD rate and the RR for each of the 32 incentivised conditions in the year 2009, prior to the start of the SDD pricing policy. Observed rates for planned conditions are highlighted in light grey, and those for emergency conditions in dark grey. There is marked heterogeneity both in terms of the observed SDD rate and the remaining gap towards the RR, i.e. the potential for growth.

Observed SDD rates may change over time due to unrelated changes in medical technology which facilitates SDD treatment for specific subpopulations of patients. To account for this, we apply an indirect standardisation approach to calculate risk-adjusted quarterly rates of SDD for each hospital and condition in our dataset, holding the relationship between patient characteristics and the probability of SDD constant over time. We construct a set of risk-adjustment variables from HES including patient age (coded as a categorical variable in 10-year bands with separate categories for 19-24 and >85), gender (male = 1), number of Elixhauser comorbidities (coded as 0, 1, 2-3, 4-6 and 7+) (Elixhauser et al. 1998) and whether the patient had any past emergency admissions within 365 days (yes = 1). As a measure of socio-economic status, we use the income deprivation score of the English Indices of Deprivation 2010 for the patients' local area of residence[11] (McLennan et al. 2011). We estimate the relationship between the vector of observed patient characteristics X_i and the probability of SDD for all patients $i = 1, \ldots, N$ treated in the financial year 2006 using the

[11] Defined as the lower layer super output area (LSOA), with an average population of approximately 1,500 individuals.

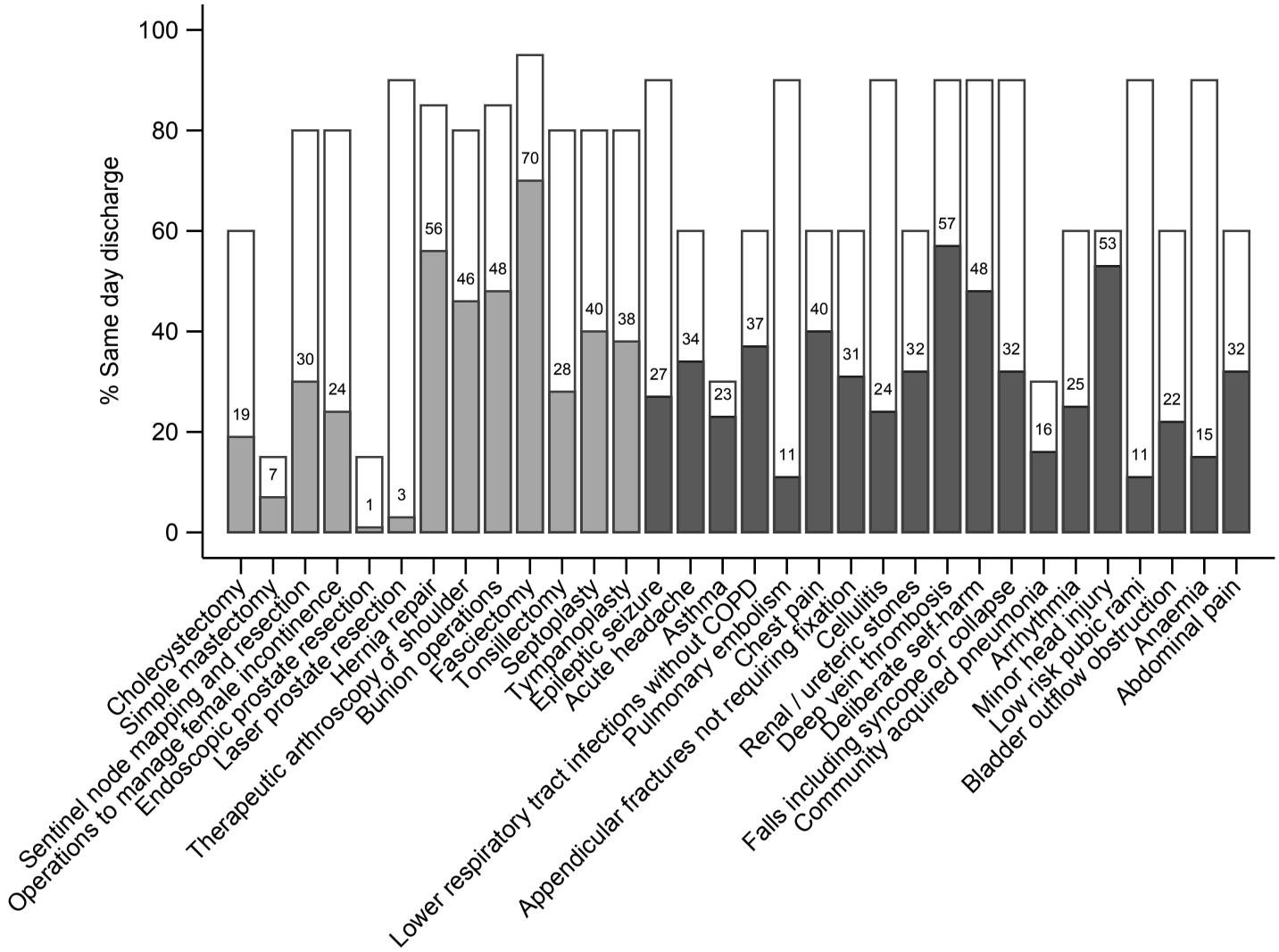


Figure 1: Observed SDD rate (grey bar) and recommended rate (white bar) for incentivised conditions in 2009

logit model[12]

$$Pr[Y_i = 1 \mid \boldsymbol{X_i}] = \frac{exp(\alpha + \boldsymbol{X_i}'\theta)}{1 - exp(\alpha + \boldsymbol{X_i}'\theta)} \tag{4}$$

where Y_i is a binary indicator that takes the value of one if the patient was admitted and discharged on the same calendar day. As our primary concern is changes in the risk relationship over time that are common to all hospitals, we do not include hospital fixed effects in this equation. The predicted probabilities $\hat{Y}_{ijt}$ for patients i in hospital j in quarter t are then used to derive the risk-adjusted hospital-quarter rate

$$\hat{Y}_{jt} = \frac{\sum_{i=1}^{N_{jt}} Y_{ijt}}{\sum_{i=1}^{N_{jt}} \hat{Y}_{ijt}} \times \bar{Y}_{2006Q2} \tag{5}$$

Equations 4 and 5 are estimated separately for each of the 191 conditions in our sample. Note that, as long as the same case-mix adjustment model is used for all periods, our choice of Quarter 2 (April-June) 2006 as the base quarter is arbitrary. Further, since the prediction model for $\hat{Y}_{jt}$ is based on large numbers of patients, we can safely ignore sampling uncertainty in parameter estimates used to adjust for case-mix differences.

Hospitals are consulted on any changes to the payment system — including the introduction of SDD bonuses applied to other conditions — approximately six months prior to the change. This gives them time to adapt to the new policy before the actual implementation, which may bias observed pre-policy rates. We therefore exclude data for the six months prior to the condition being incentivised. For some conditions eligibility criteria were refined over time to restrict the incentive to a more tightly defined patient population in which case we apply the criteria that were valid when the financial incentive first applied to ensure consistency throughout the study period.

The overall sample includes 11,336,138 patients with incentivised conditions and 21,121,500 patients with non-incentivised conditions. Descriptive statistics for case-mix variables by incentivised condition are available in Table A1 in the Appendix. Each hospital is observed for up to 34 quarters per condition. The number of hospital-quarter observations varies across the incentivised conditions and ranges from 3,022 (#5 Endoscopic prostate resection) to 9,245 (#7 Hernia repair).

[12] We use a logit regression model to avoid predicting outside the probability range of 0 to 1. This is less of an issue when drawing inference about DID regression coefficients as described in section 4.1.

4 Methods

Our empirical analysis seeks to estimate the causal effect of the SDD bonus policy on the probability that a patient admitted with an incentivised condition is discharged on the same day as admission[13]. We perform separate analyses for each of the 32 incentivised conditions. For each incentivised condition, we estimate DID and SC models, both of which aim to control for common exogenous shocks and underlying time trends by means of a comparison with a control condition. We consider as potential control conditions all non-incentivised conditions from the BADS / BAAEC directories that: (i) follow the same admission pathway (planned or emergency); (ii) have an RR ±15pp of the incentivised condition to avoid differential ceiling effects [14]; (iii) have SDD rates that are no more than 30pp apart at the start of our sample period (Q2 2006); and (iv) have at least, on average, 300 admissions per quarter over the pre-policy period.

4.1 Difference-in-difference analysis

Our DID approach relies on selecting a *single* control condition that is not affected by the SDD bonus policy but satisfies the parallel trends assumption that it responds similarly to the same external influences, for each incentivised condition. If more than one potential control condition satisfies these considerations, we select the one which minimises the difference in trends in the proportion of SDDs prior to the introduction of the pricing policy (i.e. matching on pre-trends), where pre-policy trends for each condition are estimated from separate linear regressions of $\hat{Y}_{jt}$ on a continuous measure of time as well as hospital and seasonal fixed effects.

For each incentivised condition, we then estimate the following DID model:

$$\hat{Y}_{cjt} = \beta_0 + \beta_1 SDD_c + \gamma D_t + \tau(D_t \times SDD_c) + \nu_{cj} + \varphi_{ct} + \omega_{cjt}, \tag{6}$$

where $\hat{Y}_{cjt}$ is the risk-adjusted rate of SDD in hospital j in quarter t and for condition $c \in [0, 1]$, where 1 denotes the incentivised condition, φ_{ct} is a vector of condition-specific seasonal effects (spring, summer, autumn, winter), and ν_{cj} is a vector of condition-specific hospital fixed effects,

[13] Our analysis focuses on the intensive margin. Hospitals may also respond to the financial incentive by increasing the volume of incentivised activity. However, we do not observe faster annual growths in volume of activity after the introduction of the SDD bonus (pre: 6.5% vs. post: 2.3%, p = 0.264). Furthermore, the growth in non-incentivised conditions over the 9 year period (mean = 13.3% per year) exceeds that of the incentivised conditions (mean = 5.4%). Appendix Table A2 shows annual volumes of activity for the incentivised conditions.

[14] See also Allen et al. (2016). While it is possible mathematically for SDD rates to approach 100%, we expect the RR to act as a natural ceiling that is unlikely to be breached.

which capture unobserved time-invariant differences amongst hospitals (e.g. management quality, local demand) in the propensity to discharge patients on the same day as admission[15].

The dummy variable SDD_c takes the value of 1 if condition c is incentivised by the SDD bonus and 0 otherwise and D_t is a dummy variable that takes on the value of 1 after the introduction of the SDD bonus in $t = t^\star$, and zero otherwise. The coefficient of interest is τ, which denotes the average treatment effects on the treated (ATT) over the post-policy period. ω_{cjt} is an idiosyncratic error term.

We also identify separate ATTs τ_k for each of the post-policy years $k = 1 \dots K$ by replacing the single dummy variable of D_t with a vector of dummy variables, each taking the value 1 for a specific post-policy year k. These models thus allow for a delayed impact of the SDD policy which may be because clinical processes take time to be reorganised. Alternatively, positive policy effects may fade over time due to increasing marginal costs of further improvements.

All models are estimated as linear probability models with standard errors clustered at hospital level.

4.2 Synthetic control analysis

The validity of our DID estimates may be compromised by two challenges. First, in our study, we consider a large pool of potential control conditions, several of which may be suitable to model the counterfactual outcome. The results of the DID analysis may be sensitive to the choice of control condition, for example because of idiosyncratic shocks or measurement error in the control condition. Second, while we select DID control conditions based on pre-policy trends, the assumption of parallel trends applies to unobserved counterfactual outcomes and can therefore never be tested Abadie et al. (2010). If the relationship between time-invariant unobservables and the outcome changes over time, the parallel trend assumption is violated (Imbens and Wooldridge 2009). The SC method proposed by Abadie and Gardeazabal (2003), Abadie et al. (2010), and Abadie et al. (2015) can address both of these challenges. The method constructs a synthetic control condition as a weighted combination of all potential control conditions, thus considering all relevant information in predicting the counterfactual outcome and thereby lifting reliance on a specific control condition. Furthermore, by matching on levels, the SC method provides reassurance that the synthetic control

[15] We allow for hospital fixed effects to vary between the intervention and the control condition to account for any differences in a hospital's relative propensity to discharge patients with different clinical conditions on the same day. For example, a hospital may be *5pp* more likely than the average hospital to discharge patients with the incentivised condition on the same day and *12pp* more likely to do so for patients with the control condition. In this case, forcing a common hospital fixed effect for both groups would be inappropriate.

condition is well matched to the incentivised condition on time-invariant unobservables and that both have similar scope for improvement (and, in this study, a similar risk of ceiling effects).

The SC method requires a panel data structure with the same units of observation being followed over time. We aggregate the risk-adjusted hospital-quarter data to national SDD rates at the level of condition-quarters based on hospitals' quarterly volumes of patients. The pool of potential control conditions is the same as for the DID analysis. Each potential control condition is assigned a non-negative weight (which together sum to 1) according to a loss function that minimises the discrepancy of the incentivised and SC conditions in terms of pre-policy SDD rates, expressed as the root mean squared prediction error (RMSPE), and a set of average pre-policy patient characteristics (see Section 3). The difference between observed and counterfactual outcomes provides an estimate of the ATT and can be evaluated over different time periods to recover both τ_k and τ[16].

The SC method applies a different inference framework than standard econometric analysis, which poses a challenge for comparative inference. As there is only a single observation per condition and time point it is not possible to construct traditional standard errors. Instead, we adopt the approach of placebo tests originally proposed by (Abadie et al. 2010). We estimate a set of SC models, as described above, but treat each potential control condition in turn as if it was the incentivised condition, with the incentivised condition added to the pool of potential control conditions. In each iteration, we calculate the ratio of RMSPE in the pre- and post-intervention periods. P-values are constructed as the proportion of RMPSE ratios that are at least as large as that of the original model for the incentivised condition.[17] We convert these placebo p-values to standard errors through a normal approximation. The quality of this inference framework relies on the number of potential control conditions; for example, with only 19 potential control conditions, the smallest p-value that could be calculated is $\frac{1}{1+19} = 0.05$. Note that no standard errors can be computed if $p = 1$.

All computations are performed using the user-written `synth` command in Stata 14.

[16] The estimated treatment effects are approximately unbiased under two key assumptions: a linear relationship between the covariates and the outcome variable and a sufficiently long pre-policy time period relative to the variance of the error term.

[17] Because the main estimate is also compared against itself, the numerator of this ratio is always ≥ 1 and the denominator is $V + 1$, where V is the number of potential controls.

5 Results

5.1 Model diagnostic and control group selection

Table 2 presents descriptive statistics for the 32 incentivised conditions and corresponding control conditions under the two methodologies. For each incentivised condition we calculate the pre-policy trend (i.e. linear growth per quarter) in case-mix adjusted SDD rates as well as the same information (expressed as deviations) for the control conditions, which serve as a diagnostic device of the *parallel trend* assumption of the DID method. We also calculate (differences in) pre-policy levels, which are informative about the level equivalence assumption of the SC method. Time-series graphs of SDD rates for incentivised and control conditions are presented in the online appendix.

Our two selection approaches identify control conditions that are closely matched on pre-policy trends with an average absolute deviation of 0.3pp per year for DID control conditions and 0.6pp per year for SC control conditions (Columns 3 and 5). Only one DID control condition (#19 Chest pain) shows a divergence in SDD rates of >1pp per year, suggesting that the parallel trends assumption of the DID method are generally met in our analyses. In addition, the SC control conditions are well matched in terms of levels ($|\ \overline{\Delta Level}\ |$) = 4pp) (Column 6) although this is traded off against worse fit in terms of trends, with a larger number of conditions showing divergences of >1pp per year.

Overall, for both methods and diagnostic statistics, the fit of the control condition is better for planned care, where there is a large number of potential control conditions to choose from (16 to 85), than for emergency care (2 to 7). The small number of emergency control conditions also limits the scope for inference after SC estimation. Only eleven out of 32 incentivised conditions have a set of at least 20 potential control conditions necessary to generate p-values <0.05.

5.2 Policy effect on SDD rates

Tables 3 and 4 present the results of our DID and SC analysis. Figures 2 and 3 summarise the main quantities of interest, the estimated ATT over the post-policy period (τ) and associated 95% confidence intervals, in the form of forest plots. Results are presented for all 32 conditions, with light grey, dashed confidence intervals flagging control conditions with trend divergence of >1pp per year, i.e. where we deem the underlying identification assumptions to be less clearly met.

Table 2: Comparison of levels and slopes of incentivised and control conditions

		Incentivised condition		Control group					
				DID		SC			
								Non-incentivised conditions	
#	Condition	Slope	Level	ΔSlope	ΔLevel	ΔSlope	ΔLevel	Potential	Selected (*)
1	Cholecystectomy	0.008	0.122	-0.00002	-0.0029	0.00001	-0.0028	16	4
2	Simple mastectomy	0.000	0.038	-0.00005	-0.0245	0.00021	-0.0152	20	6
3	Sentinel node mapping and resection	0.004	0.332	0.00017	0.2841	0.00274	-0.0287	66	4
4	Operations to manage female incontinence	0.008	0.119	0.00009	0.2538	0.00054	-0.0224	66	5
5	Endoscopic prostate resection	0.000	0.018	0.00002	0.0515	0.00050	0.0047	20	2
6	Laser prostate resection	0.000	0.023	-0.00015	0.0276	0.00095	-0.0152	83	1
7	Hernia repair	0.007	0.479	-0.00005	-0.0146	0.00024	0.0065	85	9
8	Therapeutic arthroscopy of shoulder	0.013	0.320	-0.00090	0.0652	0.00044	0.0414	66	4
9	Bunion operation	0.012	0.307	0.00013	0.0778	0.00083	-0.0407	85	3
10	Fasciectomy	0.011	0.569	0.00052	-0.1839	-0.00008	-0.0305	79	5
11	Tonsilectomy	0.013	0.130	0.00061	0.1882	-0.00109	-0.0391	66	2
12	Septoplasty	0.014	0.248	0.00128	0.1077	0.00014	0.0013	16	4
13	Tympanoplasty	0.011	0.229	-0.00039	0.2809	0.00018	0.0005	66	6
14	Epileptic seizure	0.002	0.202	-0.00040	0.1154	0.00039	-0.0174	7	4
15	Acute headache	0.002	0.311	-0.00031	-0.2282	-0.00247	0.0931	2	1
16	Asthma	0.002	0.200	-0.00211	-0.0951	-0.00254	0.0940	2	1
17	Lower respiratory tract infections without COPD	0.004	0.291	-0.00196	-0.2083	-0.00370	0.0809	2	1
18	Pulmonary embolism	0.003	0.080	-0.00040	0.1110	-0.00121	0.0106	7	2
19	Chest pain	0.004	0.315	-0.00260	-0.2321	-0.00448	0.1157	2	1
20	Appendicular fractures not requiring fixation	0.003	0.221	-0.00103	-0.1379	-0.00313	0.0387	2	1
21	Cellulitis	0.002	0.184	-0.00040	0.1333	0.00018	0.0058	7	3
22	Renal/ureteric stones	0.004	0.239	-0.00233	-0.1554	-0.00448	0.0258	2	1
23	Deep vein thrombosis	0.005	0.421	0.00018	-0.2422	-0.00377	0.1781	7	1
24	Deliberate self-harm	0.002	0.390	0.00023	-0.0733	0.00065	0.1164	7	1
25	Falls including syncope or collapse	0.003	0.257	0.00002	-0.0365	0.00089	0.0029	7	3
26	Community acquired pneumonia	0.002	0.102	-0.00207	0.0024	-0.00189	-0.0090	2	2
27	Arrhythmia	0.003	0.195	-0.00124	-0.1116	-0.00405	-0.0243	2	1
28	Minor head injury	0.003	0.444	-0.00013	-0.2530	-0.00185	0.1757	7	1
29	Low risk pubic rami	0.000	0.071	0.00126	0.2464	0.00115	0.0193	7	2
30	Bladder outflow obstruction	0.003	0.171	-0.00133	-0.0877	-0.00399	-0.0561	2	2
31	Anaemia	0.001	0.083	0.00006	0.2339	0.00037	0.0155	7	3
32	Abdominal pain	0.002	0.238	0.00003	-0.1550	-0.00275	0.0510	2	1

Notes: Slope and Level indicate the trend in SDD rates and the average SDD rate in the intervention group prior to the policy introduction. ΔSlope and ΔLevel denote the difference between intervention and control groups. All slope estimates are per quarter.

* *Number of potential control conditions with a non-zero weight assigned to them.*

Table 3: Average treatment effect on the treated - DID analyses

		τ post-policy ATT		τ_1 (Year 1 ATT)		τ_2 (Year 2 ATT)		τ_3 (Year 3 ATT)		τ_4 (Year 4 ATT)		τ_5 (Year 5 ATT)	
#	Condition	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
1	Cholecystectomy	0.104**	0.033	0.065 *	0.028	0.090 *	0.036	0.111 **	0.042	0.113 **	0.041	0.127 **	0.045
2	Simple mastectomy	0.084***	0.025	0.022	0.017	0.087 **	0.028	0.130 ***	0.035	0.096 **	0.034		
3	Sentinel node mapping and resection	0.201***	0.024	0.102 ***	0.023	0.188 ***	0.026	0.244 ***	0.028	0.252 ***	0.030		
4	Operations to manage female incontinence	0.075**	0.027	0.044	0.025	0.079 **	0.028	0.106 ***	0.032	0.083 *	0.033		
5	Endoscopic prostate resection	0.005	0.006	0.005	0.007	-0.002	0.008	0.010	0.008	0.008	0.010		
6	Laser prostate resection	-0.016	0.009	-0.015	0.009	-0.023 *	0.011	-0.016	0.012	-0.011	0.012		
7	Hernia repair	0.019*	0.008	0.020 *	0.008	0.024 **	0.009	0.011	0.010	0.020	0.011		
8	Therapeutic arthroscopy of shoulder	0.030	0.028	-0.013	0.027	0.024	0.032	0.040	0.036	0.071 *	0.034		
9	Bunion operation	0.036	0.024	-0.008	0.025	0.025	0.029	0.048	0.030	0.091 **	0.031		
10	Fasciectomy	0.036	0.027	0.015	0.027	0.031	0.030	0.046	0.033	0.054	0.033		
11	Tonsillectomy	0.109***	0.024	0.065 **	0.023	0.105 ***	0.028	0.153 ***	0.027				
12	Septoplasty	0.085*	0.034	0.045	0.030	0.099 **	0.038	0.114 **	0.037				
13	Tympanoplasty	0.084***	0.015	0.073 ***	0.016	0.095 ***	0.018						
14	Epileptic seizure	0.008	0.007	0.018	0.010	0.008	0.009	-0.002	0.010				
15	Acute headache	0.001	0.005	-0.007	0.005	0.003	0.006	0.006	0.006				
16	Asthma	-0.009	0.009	-0.002	0.012	-0.016	0.012	-0.012	0.013				
17	Lower respiratory tract infections without COPD	0.043***	0.008	0.028 **	0.010	0.051 ***	0.010	0.055 ***	0.010				
18	Pulmonary embolism	0.035***	0.009	0.015	0.011	0.020	0.012	0.067 ***	0.014				
19	Chest pain	0.060***	0.005	0.043 ***	0.005	0.062 ***	0.005	0.078 ***	0.006				
20	Appendicular fractures not requiring fixation	-0.018	0.010	-0.009	0.010	-0.017	0.012	-0.028 *	0.011				
21	Cellulitis	0.063***	0.011	0.035 **	0.013	0.068 ***	0.014	0.083 ***	0.016				
22	Renal/ureteric stones	0.045***	0.009	0.034 ***	0.009	0.046 ***	0.010	0.055 ***	0.011				
23	Deep vein thrombosis	-0.024	0.016	-0.054 ***	0.016	-0.031	0.018	0.012	0.019				
24	Deliberate self-harm	-0.035***	0.007	-0.018	0.010	-0.034 ***	0.010	-0.056 ***	0.010				
25	Falls including syncope or collapse	0.002	0.004	-0.009	0.005	0.007	0.005	0.010	0.006				
26	Pneumonia	0.008	0.008	0.000	0.011	0.015	0.011						
27	Arrhythmia	0.004	0.003	0.007	0.004	0.001	0.004						
28	Minor head injury	-0.007	0.011	-0.019	0.013	0.006	0.014						
29	Low risk pubic rami	-0.018*	0.008	-0.006	0.011	-0.032 **	0.010						
30	Bladder outflow obstruction	0.000	0.010	0.007	0.011	-0.006	0.012						
31	Anaemia	0.042***	0.010	0.034 **	0.011	0.047 ***	0.013						
32	Abdominal pain	-0.008	0.006	-0.010	0.006	-0.006	0.007						
	Pooled - planned	0.051***	0.006										
	Pooled - emergency	0.014***	0.002										

*** p<0.001; ** p<0.01; * p<0.05
Standard errors (SEs) are clustered at hospital level.
ATT = Average treatment effect on the treated.

Table 4: Average treatment effect on the treated - SC analyses

#	Condition	Potential control conditions	τ (post-policy ATT)		τ_1 (Year 1 ATT)		τ_2 (Year 2 ATT)		τ_3 (Year 3 ATT)		τ (Year 4 ATT)		τ (Year 5 ATT)	
			Est	p-value	Est	p-value	Est	p-value	Est	p-value	Est	p-value	Est	p-value
1	Cholecystectomy	16	0.097	0.059	0.064	0.059	0.096	0.059	0.136	0.059	0.180	0.059	0.204	0.059
2	Simple mastectomy	20	0.026	0.095	0.006	0.095	0.065	0.095	0.099	0.095	0.055	0.190		
3	Sentinel node mapping and resection	66	0.091	0.030	0.060	0.015	0.129	0.015	0.191	0.015	0.199	0.015		
4	Operations to manage female incontinence	66	0.018	0.104	0.012	0.313	0.039	0.134	0.044	0.164	0.057	0.075		
5	Endoscopic prostate resection	20	0.004	0.286	0.010	0.286	0.005	0.714	0.011	0.381	0.008	0.905		
6	Laser prostate resection	83	-0.021	0.119	-0.038	0.012	-0.039	0.095	-0.049	0.143	-0.038	0.262		
7	Hernia repair	85	0.003	0.628	0.001	0.930	0.009	0.500	0.006	0.709	0.008	0.849		
8	Therapeutic arthroscopy of shoulder	66	0.011	0.985	0.024	0.343	0.035	0.284	0.000	0.955	0.001	0.955		
9	Bunion operation	85	0.001	0.919	0.015	0.453	0.031	0.163	-0.020	0.628	-0.019	0.779		
10	Fasciectomy	79	0.022	0.088	0.041	0.013	0.073	0.013	0.043	0.175	0.047	0.175		
11	Tonsillectomy	66	0.025	0.134	0.051	0.030	0.060	0.060	0.074	0.075				
12	Septoplasty	16	0.015	0.059	0.037	0.176	0.049	0.176	0.031	0.353				
13	Tympanoplasty	66	0.006	0.104	0.025	0.090	0.031	0.149						
14	Epileptic seizure	7	-0.005	0.125	0.001	0.375	-0.014	0.625	-0.026	0.125				
15	Acute headache	2	0.053	0.667	0.121	0.333	0.139	0.333	0.148	0.333				
16	Asthma	2	0.030	1.000	0.096	0.667	0.079	0.667	0.083	0.667				
17	Lower respiratory tract infections without COPD	2	0.053	0.333	0.159	0.333	0.190	0.333	0.205	0.333				
18	Pulmonary embolism	7	0.013	0.250	-0.004	1.000	0.033	0.500	0.067	0.125				
19	Chest pain	2	0.089	0.667	0.229	0.333	0.256	0.333	0.278	0.333				
20	Appendicular fractures not requiring fixation	2	0.028	0.333	0.084	0.333	0.084	0.333	0.080	0.333				
21	Cellulitis	7	0.001	0.500	-0.010	0.750	-0.011	0.750	0.032	0.375				
22	Renal/ureteric stones	2	0.051	0.333	0.129	0.333	0.147	0.333	0.165	0.333				
23	Deep vein thrombosis	7	0.086	0.625	0.226	0.125	0.271	0.125	0.312	0.125				
24	Deliberate self-harm	7	0.034	1.000	0.120	0.125	0.107	0.500	0.083	0.500				
25	Falls including syncope or collapse	7	0.007	0.375	0.014	0.625	0.027	0.500	0.030	0.375				
26	Community acquired pneumonia	2	0.008	0.667	0.032	0.667	0.047	0.667						
27	Arrhythmia	2	0.012	0.333	0.049	0.333	0.049	0.333						
28	Minor head injury	7	0.048	0.500	0.285	0.125	0.290	0.125						
29	Low risk pubic rami	7	-0.017	0.125	-0.053	0.125	-0.076	0.125						
30	Bladder outflow obstruction	2	0.007	0.333	0.047	0.333	0.038	0.667						
31	Anaemia	7	0.002	0.625	0.002	0.875	0.015	0.500						
32	Abdominal pain	2	0.025	0.667	0.099	0.333	0.112	0.333						
	Pooled - planned		0.033											
	Pooled - emergency		0.127											

Note: The reported p-values are derived from the result of placebo tests using a normal approximation. SE are not estimated as part of the SC routine and are therefore not reported. The range of possible p-values for each condition is constraint by the number of potential control groups.
ATT = Average treatment effect on the treated.

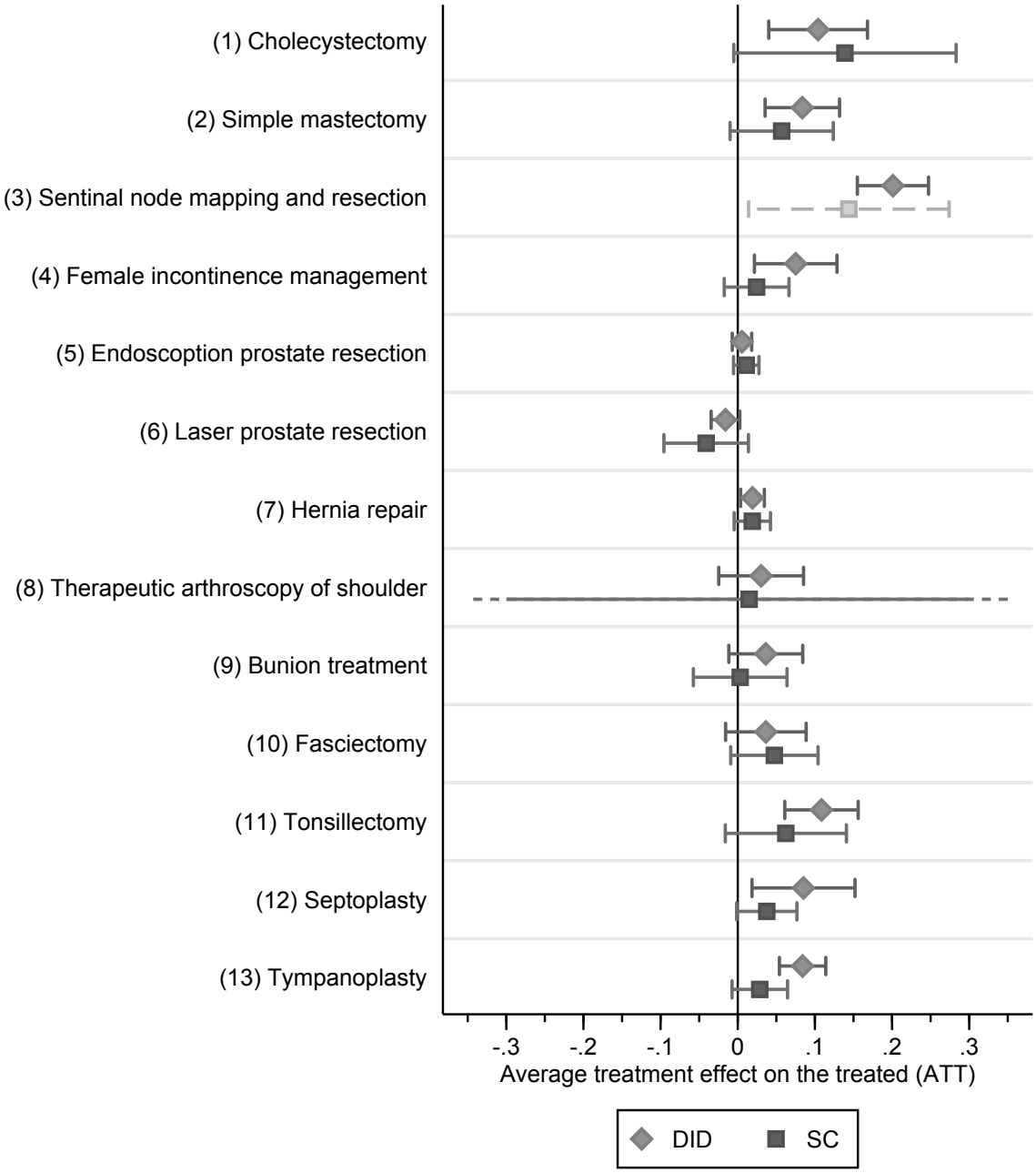


Figure 2: Average change in SDD rate over post-policy - planned conditions

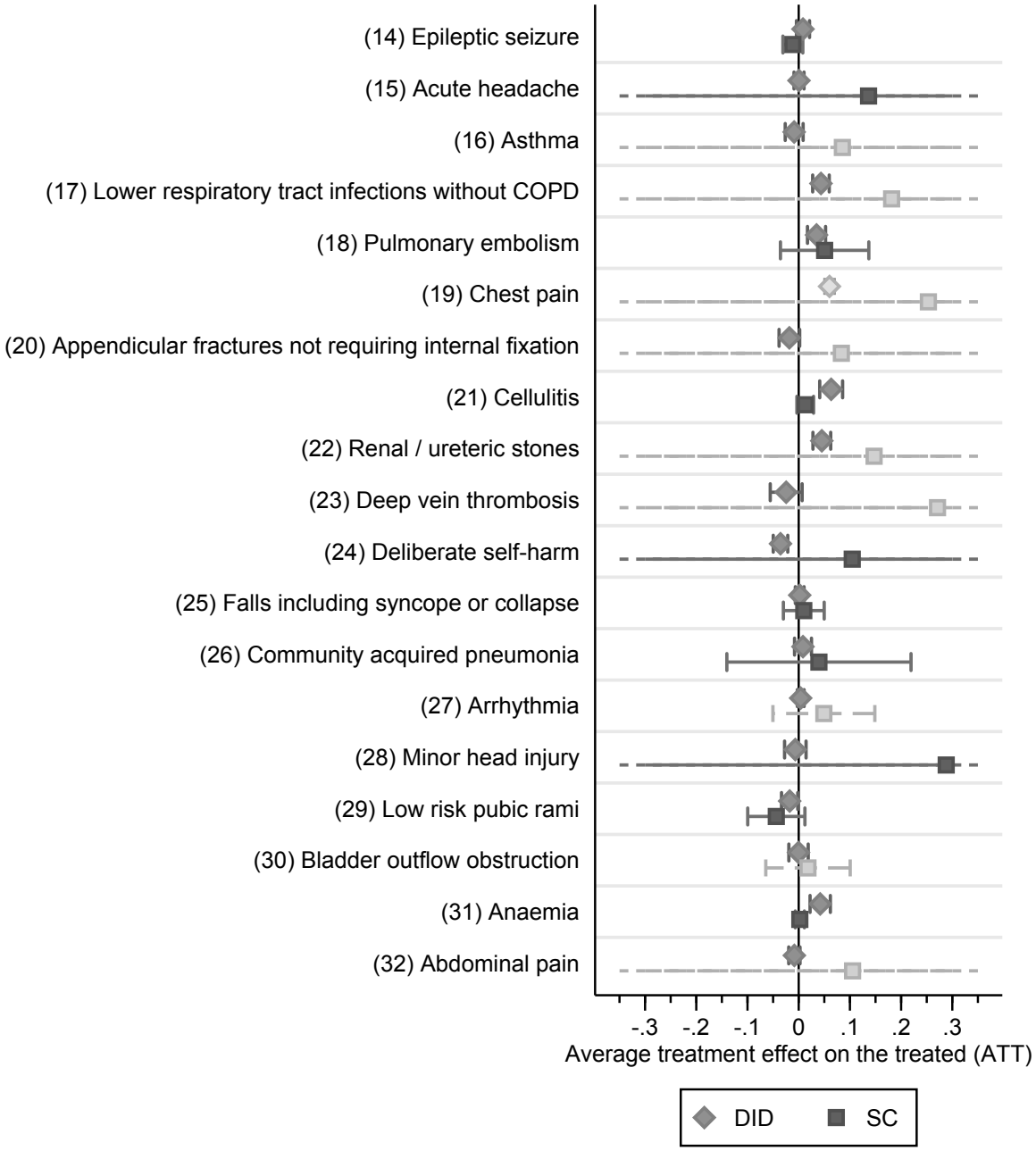


Figure 3: Average change in SDD rate over post-policy - emergency conditions

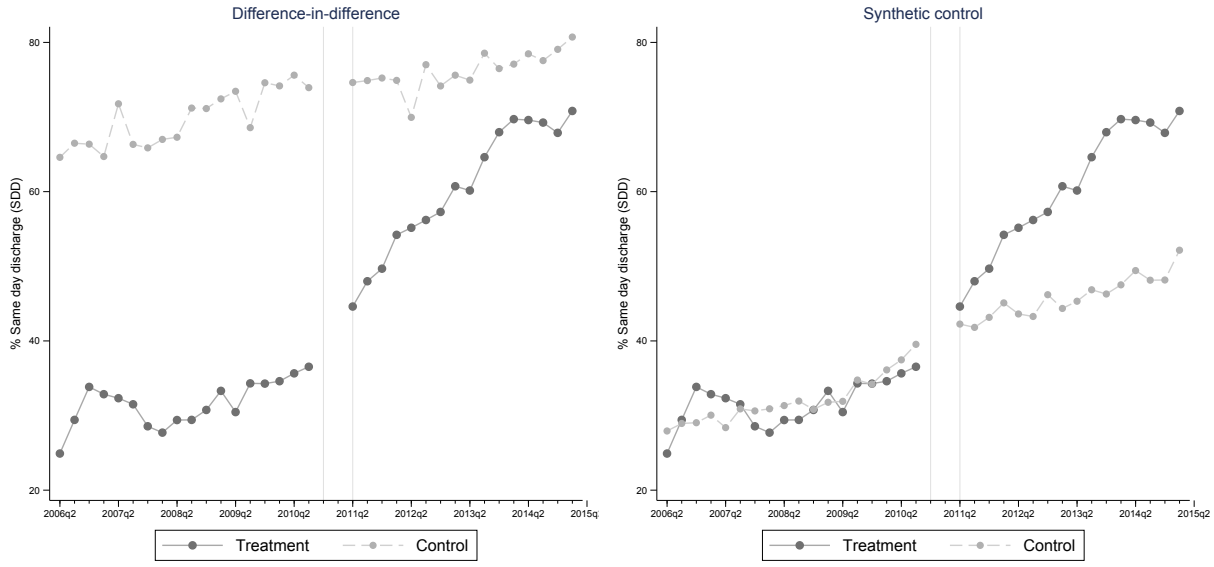


Figure 4: Time trends for incentivised and control conditions - Sentinel node mapping and resection

For the planned conditions, the results of DID analysis suggest that the policy led to a statistically significant increase in SDD rates for 8 of the 13 incentivised conditions.[18]. The estimated policy effects are heterogeneous in size, ranging from -1.6pp to 21.7pp, with three instances of more than 10pp. However, the results of the SC analysis call for a more conservative interpretation. Although the point estimates under both methods are typically quite similar, the confidence intervals around the SC estimates are substantially wider, even in instances where a large number of potential control conditions exist. As a result, there is only planned condition (#3 Sentinal node mapping) where a statistically significant increase in SDD rates can be ascribed to the policy. This is shown as an example in Figure 4.

For emergency conditions, the DID analysis identifies statistically significant positive effects for six conditions and negative effects for two conditions. The size of the effects is generally smaller than those estimated for planned conditions, with no point estimate exceeding 6pp. Given the small number of potential control conditions, the SC estimates are less reliable and deviate substantially from the DID results. Moreover, the placebo tests cannot reject the possibility that these results reflect chance variation, as evidenced by very wide confidence intervals.

The pooled effect across conditions according to our DID results are a 5.3pp increase in the probability of SDD for planned patients, and a 1.4pp increase for emergency patients (Tables 3

[18] The number of incentivised conditions with statistically significant DID estimates reduces to 12 (five planned conditions and 7 emergency conditions) after applying the Bonferroni adjustment for multiple comparisons. None of the SC analyses yields statistically significant results.

and 4), both of which are statistically significant at p<0.001.[19] These DID results translate into approximately 28,400 additional patients (95% CI: 23,297 to 33,502) admitted, treated and discharged on the same day in a year across all incentivised conditions (Figure 5).[20] Most of these additional patients receive treatment for chest pain, where a small change in SDD rates applies to a large patient population.

Figure 6 plots out the development of the policy effects for each of the 32 incentivised conditions over time based on the DID model with interactions. The estimated developments are generally non-linear, with some conditions experiencing an immediate response to the change in financial incentives and subsequent flattening out, whereas others show a slow increase in SDD over time. There is no single pattern to these developments with all possible permutations present.

5.3 Robustness checks

We conduct two robustness checks to rule out alternative explanations of our results which are presented in Table 5. First, the introduction of incentives to increase SDD rates for some conditions might lead to changes in SDD provision more broadly. These spillovers might be positive, for example if clinicians apply their new skills to non-incentivised clinical conditions, or negative, for example if increasing the provision of SDD care requires resources which might be in demand for other patients, such as specialised day surgery beds. Spillover effects are most likely to occur within the same clinical department, as departments are where hospital resources such as clinical personal and beds are managed on a day-to-day basis. To test for spillovers, we re-estimate our analyses excluding potential control conditions that are performed in the same clinical department as the incentivised condition.[21] We find our results to be substantively unchanged, suggesting that spillovers are unlikely to drive our main estimates.

Second, for planned conditions, hospitals only receive the higher SDD price if they both schedule and provide SDD care. Hospitals that are already achieving high SDD rates prior to the policy but record poorly whether they have scheduled that care in advance to be delivered on the same day, may therefore be able to increase their payment simply by better recording scheduling plans. If so, observed changes in the incentivised outcome may not reflect changes in patient care but

[19] The overall effects are calculated as weighted averages, where the weights (w_m) are given by the size of the patient population for each incentivised condition $m = 1 \dots M$ divided by the size of the patient population overall. The corresponding standard errors are calculated as $\sqrt{\sum_m^M SE_m^2 \times w_m^2}$.

[20] The additional patients treated as SDD across all incentivised conditions in a given year is $\sum_{cm=1}^{32} \tau \bar{N}_{cm} \sum_{m=1}^{32} \tau \bar{N}_m$ where $\bar{N}_m$ is the number of patients within the scope of each incentivised condition m in the average post-policy year.

[21] All emergency conditions are considered to be part of the same specialty of Emergency Medicine.

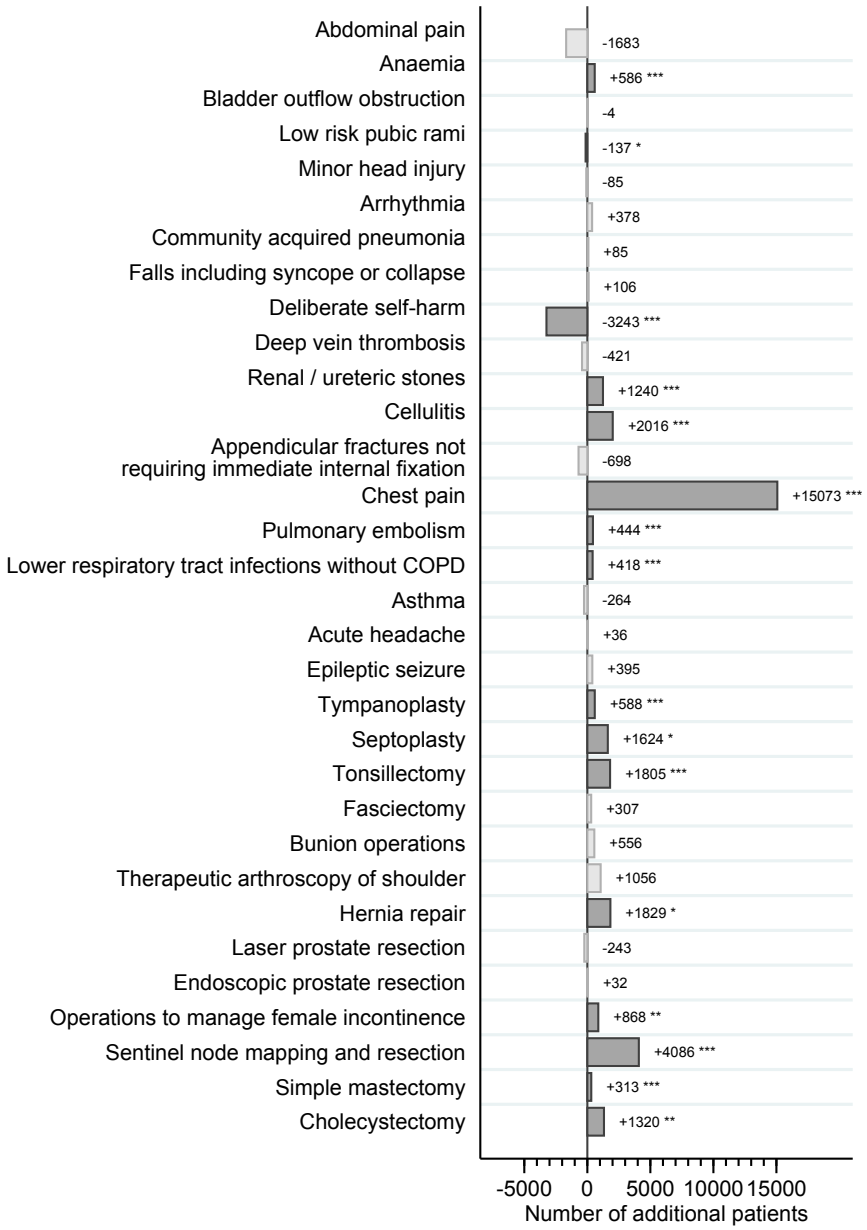


Figure 5: Additional SDD patients per year based on ITS estimates

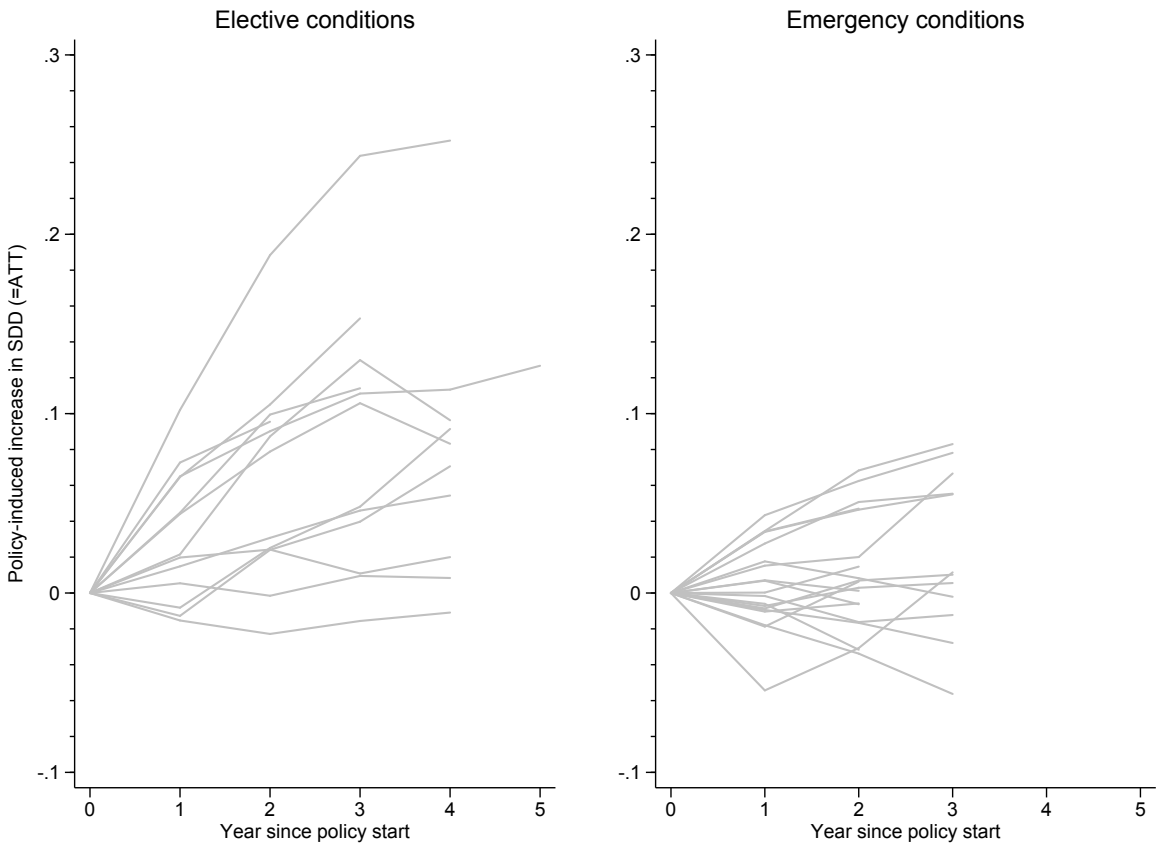


Figure 6: ATTs over time

Table 5: Sensitivity analyses

	Difference-in-Difference						Synthetic control								
	Main model		Control conditions from different departments		LoS = 0		Main model			Control conditions from different departments			LoS = 0		
BPT	ATT	SE	ATT	SE	ATT	SE	ATT	Placebo tests	p-value	ATT	Placebo tests	p-value	ATT	Placebo tests	p-value
1	0.104***	0.033	0.104***	0.033	0.073***	0.025	0.139	16	0.059	0.167	14	0.067	0.158	16	0.059
2	0.084***	0.025	0.084***	0.025	0.103***	0.029	0.057	20	0.095	0.065	19	0.050	0.072	20	0.048
3	0.201***	0.024	0.201***	0.024	0.221***	0.018	0.144	66	0.030	0.147	63	0.031	0.136	66	0.045
4	0.075***	0.027	0.075***	0.027	0.111***	0.015	0.024	66	0.254	0.017	59	0.533	0.028	66	0.254
5	0.005	0.006	0.005	0.006	0.019***	0.009	0.011	20	0.190	0.010	18	0.158	0.011	20	0.381
6	-0.016	0.009	-0.016	0.009	-0.092***	0.012	-0.041	83	0.143	-0.041	67	0.074	-0.212	83	0.321
7	0.019***	0.008	0.019***	0.008	0.031***	0.006	0.019	85	0.116	0.018	78	0.165	0.006	85	0.407
8	0.030	0.028	0.030	0.028	-0.060***	0.017	0.015	66	0.955	0.014	59	0.950	0.043	66	0.567
9	0.036	0.024	0.036	0.024	0.037***	0.018	0.003	85	0.919	0.003	75	0.908	0.013	85	0.826
10	0.036	0.027	0.036	0.027	0.046***	0.013	0.047	79	0.100	0.046	70	0.113	0.039	79	0.063
11	0.109***	0.024	0.109***	0.024	0.049***	0.016	0.062	66	0.119	0.102	61	0.048	0.069	66	0.119
12	0.085***	0.034	0.085***	0.034	0.037	0.031	0.038	16	0.059	0.070	13	0.071	0.058	16	0.059
13	0.084***	0.015	0.084***	0.015	0.098***	0.020	0.029	66	0.119	0.029	66	0.119	0.009	66	0.567

*** p<0.001; ** p<0.01; * p<0.05
BPT = Best Practice Tariff, ATT = Average treatment effect on the treated, SE = Standard error, LoS = Length of Stay
Standard errors in difference-in-difference analysis are clustered at hospital level.

just coding practice. We therefore also estimate models where the dependent variable is a simple indicator of SDD (i.e. LoS=0), independent of scheduling. Our findings are broadly similar across DID analyses. In general, policy effects on LoS=0 rates are larger than those based on meeting the exact conditions for the SDD bonus, suggesting our main analysis is conservative in measuring the impact of the policy on patient care, as hospitals did not always plan or report the planning of SDD despite carrying it out. Exceptions to this general finding are conditions #1 Cholecystectomy, #11 Tonsilectomy and #12 Septoplasty, where effects on LoS=0 are smaller but still positive. Furthermore, for conditions #6 Laser prostate resection and #8 Therapeutic arthroscopy of shoulder our LoS=0 estimates indicate large negative effects of the policy which are also significant. Comparisons of SC analyses indicate generally similar magnitudes of policy effects.

5.4 Association with incentive design features

Thus far, our results have demonstrated that the response to the SDD bonus policy varies substantially across incentivised conditions. We now investigate if this variation is associated with features of the design of SDD incentives. Since the 32 conditions incentivised by the policy vary in the size of the price differential $P_1^{SDD} - P_1^{ON}$ relative to the base price P_1^{ON}, we compute the elasticities of the policy response with respect to price changes as

$$\epsilon = \frac{\tau / \bar{Y}_{Pre}}{(P_1^{SDD} - P_1^{ON})/P_1^{ON}} \tag{7}$$

where $\bar{Y}_{Pre}$ is the observed outcome for the incentivised condition in the year before the announcement period. Focussing on the DID estimates, we find a median elasticity of 0.24 across the 13 planned conditions, and 0.01 across the 19 emergency conditions. Five conditions show an elasticity above 1.

As there are just 32 conditions, it is not possible to conduct multivariate regression analysis of incentive design features that may affect the elasticity of the policy response. We therefore resort to univariate correlation analyses which are presented in the form of scatter plots in Figure 7. Hospitals may respond more strongly for conditions offering relatively higher financial returns. Figures 7a and 7b plot the elasticities as a function of the post-policy SDD price P_1^{SDD} and as a function of the price difference $P_1^{SDD} - P_1^{ON}$. Figure 7c shows the association between the policy response and the total incentive, capturing both price and cost differences between SDD and ON, the latter being approximated by information on average costs in the year prior to the policy introduction.

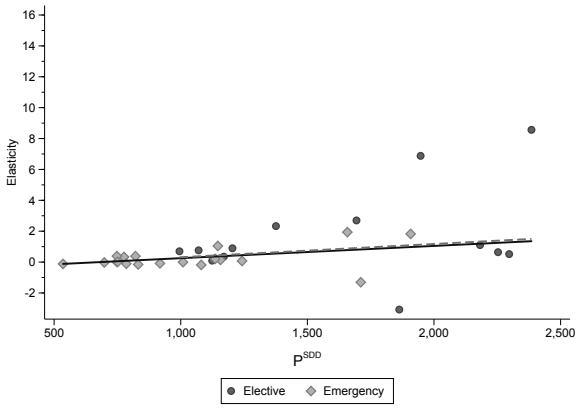
We find suggestive evidence that larger elasticities are concentrated in conditions with higher SDD prices, but not with larger price differences. Moreover, elasticities appear to increase in the size of the total incentive $\Delta(P - AC) = (P_1^{SDD} - AC_0^{SDD}) - (P_1^{ON} - AC_0^{ON})$ but only for planned SDD conditions.

We also explore whether responses appear to be driven by clinical reasons. We hypothesise that responses to the SDD bonus are more pronounced if SDD pre-policy rates are lower and the gap to the RR is higher, therefore giving more scope for improvement. Figure 7d provides some support that larger elasticities occur for planned conditions with lower pre-policy SDD rates. However, somewhat counterintuitively, Figure 7e suggests a negative relationship between the elasticities and the gap between existing practice (i.e. pre-policy SDD rate) for planned SDD care. One potential mechanism for this finding is that the size of gap between existing practice and recommended rate is larger when the costs or other limitations to higher SDD rates discussed above are larger. In such cases, the additional incentive created by the policy may still be insufficient for a larger number of hospitals, reflected in a lower national response.

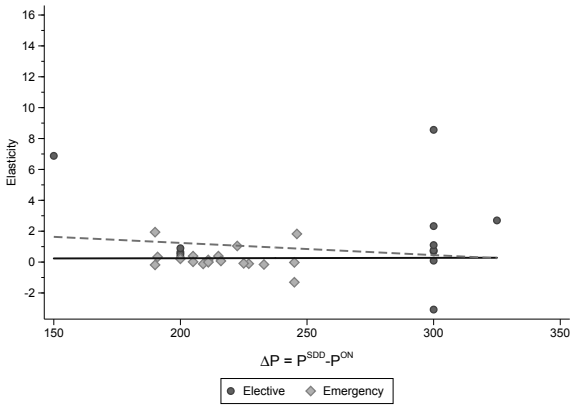
6 Conclusions

We have assessed the long-term impact of a generous pricing policy designed to encourage hospitals to treat patients as a 'same day discharge', involving admission, treatment and discharge on the same calendar day. Despite being considered clinically appropriate and having lower costs, English policy makers have been frustrated by the low rates of SDD for many conditions. Consequently, in order to encourage behavioural change by doctors and hospitals, policy makers have set prices for SDD that are well above average costs and are also higher than the price for patients allocated to the same DRG who have an overnight stay.

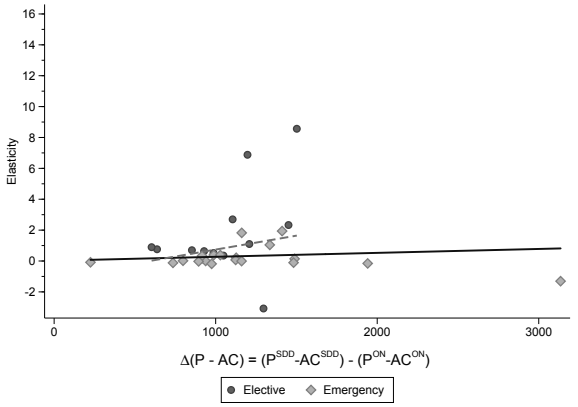
Economic theory predicts that a significant price differential would result in greater provision of treatment on an SDD basis. An early study into the policy impact for one condition, cholecystectomy, suggested that the SDD pricing policy met short-term policy objectives (Allen et al. 2016). Since this study, the policy has been rolled out to 31 more conditions. Our study set out to assess how far these earlier findings would be generalisable to these other conditions, whether short-term impacts would hold over the longer-term and what design features of the policy might explain the magnitude of any response. Based on the results of our DID analysis, we find a positive policy response for 14 of the 32 incentivised conditions, translating into approximately 28,400 more patients treated on an



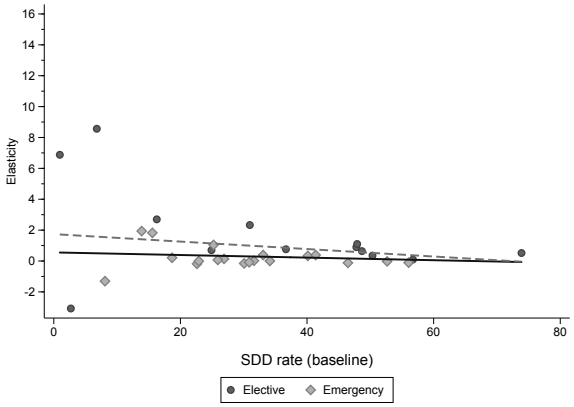
(a) SDD price



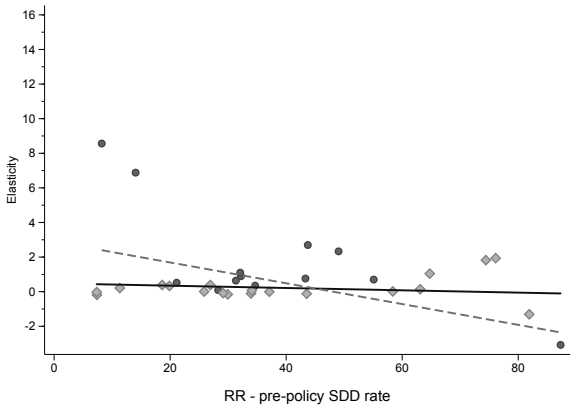
(b) % price difference SDD vs. ON



(c) Cost and price incentive



(d) Baseline SDD rate



(e) Gap between RR and baseline SDD rate

Notes: Solid line shows fitted relationship for all incentivised conditions. Dashed line shows relationship for planned conditions only.

Figure 7: Association between price elasticity of SDD care and incentive design factors

SDD basis per year. However, perhaps surprisingly, we do not find a consistent positive response across all incentivised conditions. Indeed, for two conditions the response is negative: despite the enhanced price advantage, fewer SDD treatments are provided post-policy than predicted. For others there is no apparent response. Nor are we able to identify any general temporal pattern in the policy response, with both rapid and delayed uptake of SDD practices being observed. These mixed results mirror those of the literature on P4P, which provides inconclusive evidence for the effectiveness of using financial incentives to drive quality (Milstein and Schreyögg 2016).

This lack of generalisability cautions against drawing firm conclusions from a single analysis. Indeed, cholecystectomy turns out to be the condition exhibiting the second greatest positive response among the 32 conditions. Moreover, while Milstein and Schreyögg (2016) suggested that P4P arrangements are most appropriate for emergency care, where hospitals have less opportunity to select patients, we find that the SDD pricing policy was more effective for planned care (median elasticity = 0.24) than emergency care (median = 0.01). This may be because clinicians may have ethical concerns about discharging patients in urgent need of care without a period of observation, whereas such concerns are less prominent when care is scheduled in advance. Also, emergency admissions occur at unpredictable points in the day, making it difficult to achieve SDD for some patients; particularly those admitted late in the evening. This may limit the scope for rapid increases in SDD rates in emergency conditions compared to planned conditions.

It has been argued that the limited impact of P4P schemes is due to incentives being too small (Milstein and Schreyögg 2016). In this study, for all conditions, the price incentive was more high-powered than that typically associated with P4P schemes. But there was significant variation across the conditions in terms of the relative size of the incentive, and we exploit this to investigate the association of incentive size and the estimated clinical response across 32 conditions. There is suggestive evidence that the response to the incentive was greater for conditions with higher SDD prices post policy and with lower SDD rates pre policy. There does not appear to be an association between the size of the price differential, i.e. the marginal reimbursement that hospitals attract from adopting SDD care, and the size of the response. However, there is a positive association, especially for planned conditions, when both price and cost advantages of SDD care are taken into consideration.

On the methodological side, our study highlights an important shortcoming of the SC method compared to more traditional DID analysis in a policy evaluation context commonly encountered by applied health economists. Because the SC method aims to make inference about a treatment based

on a single treated unit followed over time, the scope for statistical inference is limited to placebo tests. The quality of inference is thus dependent on the number of potential control conditions over which these placebo tests can be conducted. Even for planned SDD conditions, where there are as many as 85 potential control conditions, we only found one statistically significant result at the usual 5% critical level; compared to eight in DID analysis. This is not due to fundamentally different findings about the effectiveness of the SDD pricing policy, as point estimates were generally similar for both methods. The literature on statistical inference techniques for SC methods is rapidly evolving but has not yet reached a consensus on statistical testing (Firpo and Possebom 2018; Hahn and Shi 2017). Until then, analysts should remain cautious about drawing conclusions about policy interventions based on traditional inference thresholds, or interpret SC results as robustness checks for more traditional causal inference methods such as DID.

There are two important limitations to our study that should be addressed by future research. First, while we do not find evidence of spillovers from incentivised to non-incentivised SDD conditions, we cannot rule out that spillovers among the 32 incentivised conditions contribute to the limited overall policy effect that we observe. For example, hospitals may find it difficult to increase SDD rates for a condition that starts to be incentivised if dedicated inputs (e.g. day beds on specialised wards) are limited and have already been allocated to another condition where the incentive has been in place for longer. Our analysis treats all 32 incentivised conditions as independent and therefore cannot detect such spillovers. To address this, future research would need to develop a more complex model of inter-hospital allocation of resources that also incorporates the changes in incentive structure over time, which goes beyond the scope of the current paper. Second, our analysis focusses on changes in discharge behaviour and does not analyse effects on patients' health outcomes. The assumed welfare effects of the SDD policy are predicated upon the clinical consensus and existing evidence (e.g. Gilliard et al. (2006), Marla and Stallard (2009), Vaughan et al. (2013), and NICE (2014)) that SDD care is as safe and effective as care involving overnight stays. Future research should seek to confirm this assumption.

In conclusion, we find some evidence that hospitals respond to price signals and that payers, therefore, can use pricing instruments to improve technical efficiency. However, there appears to be substantial variation in hospitals' reactions even among similar types of financial incentives that is not explained by the size of the financial incentive or the clinical setting in which it is applied. It has been said that a randomised controlled trial demonstrates only that something works for one group of patients in one particular context but may not be generalisable (Rothwell 2005). Similarly,

a pricing policy that appears to work as intended in one area may not be effective when applied elsewhere, hence the need for continued experimentation and evaluation.

Acknowledgements

We are grateful for comments and suggestions from two anonymous referees, the editor, Hugh McLeod, Kian Chin, Paul Twigg, Simon Reif, as well as those received during presentations at the 2016 Winter Health Economists' Study Group meeting, the 3rd German Health Econometrics Workshop, the 39th Spanish Health Economics Congress, the 2017 HEB-HERO Health Economics Workshop and during seminars at NHS England, the RWI-Leibniz Institute and the Universities of Lancaster, Hamburg and Duisburg-Essen. This work reported here is based on independent research commissioned and funded by the National Institute for Health Research (NIHR) Policy Research Programme (Policy Research Unit in the Economics of Health and Social Care Systems (Ref 103/0001)). The views expressed in the publication are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care and its arm's length bodies, or other government departments. Hospital Episode Statistics are copyright ©2006-2014, re-used with the permission of NHS Digital (DSA NIC-84254-J2G1Q). All rights reserved.

References

Abadie, A., A. Diamond, and J. Hainmueller (2010). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program". *Journal of the American Statistical Association*, 105: 493–505.

— (2015). "Comparative Politics and the Synthetic Control Method". *American Journal of Political Science*, 59: 495–510.

Abadie, A. and J. Gardeazabal (2003). "The Economic Costs of Conflict: A Case Study of the Basque Country". *American Economic Review*, 93: 113–132.

Acemoglu, D., T. Hassan, and A. Tahoun (2017). "The power of the street: Evidence from Egypt's Arab spring". *The Review of Financial Studies*, 31: 1–42.

Allen, T., E. Fichera, and M. Sutton (2016). "Can Payers Use Prices to Improve Quality? Evidence from English Hospitals". *Health Economics*, 25: 56–70.

Barofsky, J., T. D. Anekwe, and C. Chase (2015). "Malaria eradication and economic outcomes in sub-Saharan Africa: Evidence from Uganda". *Journal of Health Economics*, 44: 118–136.

Bharadwaj, P., J. Johnson, and K. Løken (2014). "Smoking bans, maternal smoking and birth outcomes". *Journal of Public Economics*, 115: 72–93.

Bilgel, F. and B. Galle (2015). "Financial incentives for kidney donation: A comparative case study using synthetic controls". *Journal of Health Economics*, 43: 103–117.

Billmeier, A. and T. Nannicini (2013). "Assessing economic liberalization episodes: A synthetic control approach". *Review of Economics and Statistics*, 95: 983–1001.

Brekke, K. R., L. Siciliani, and O. R. Straume (2015). "Hospital Competition with Soft Budgets". *The Scandinavian Journal of Economics*, 117: 1019–1048.

British Association for Ambulatory Emergency Care (2014). *Directory of Ambulatory Emergency Care for Adults*. Vol. 4. Ambulatory Emergency Care Network.

British Association of Day Surgery (2006). *BADS Directory of Procedures*. 3rd ed. British Association of Day Surgery.

Cashin, C., Y.-L. Chi, P. Smith, M. Borowitz, and S. Thomson, eds. (2014). *Paying for performance in healthcare: implications for health system performance and accountability*. Buckingham: Open University Press.

Chandra, A., D. Cutler, and Z. Song (2011). "Who Ordered That? The Economics of Treatment Choices in Medical Care". In: *Handbook of Health Economics*. Ed. by M. Pauly, T. McGuire, and P. Barros. Amsterdam: Elsevier, 397–432.

Dafny, L. S. (2005). "How Do Hospitals Respond to Price Changes?" *American Economic Review*, 95: 1525–1547.

Department of Health (2002). *Implementing the new system of financial flows - payment by results: Technical guidance 2003/4*.

— (2007). *Options for the future of payment by results: 2008/09 to 2010/11*.

— (2009). *Payment by results guidance for 2010-11*.

Elixhauser, A., C. Steiner, D. Harris, and R. Coffey (1998). "Comorbidity measures for use with administrative data". *Medical Care*, 36: 8–27.

Ellis, R. and T. McGuire (1986). "Provider Behaviour under Prospective Reimbursement: Cost Sharing and Supply". *Journal of Health Economics*, 5: 129–151.

Farrar, S., D. Yi, M. Sutton, M. Chalkley, J. Sussex, and A. Scott (2009). "Has payment by results affected the way that English hospitals provide care? Difference-in-differences analysis". *British Medical Journal*, 339: b3047.

Firpo, S. and V. Possebom (2018). "Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets". *Journal of Causal Inference*, 6:

Fletcher, J. M., D. E. Frisvold, and N. Tefft (2015). "Non-Linear Effects of Soda Taxes on Consumption and Weight Outcomes". *Health Economics*, 24: 566–582.

Gilliard, N., Y. Eggli, and P. Halfon (2006). "A methodology to estimate the potential to move inpatient to one day surgery". *BMC Health Services Research*, 6: 78.

Grašič, K., A. Mason, and A. Street (2015). "Paying for the quantity and quality of hospital care: the foundations and evolution of payment policy in England." *Health Economics Review*, 5: 15.

Green, C., J. Heywood, and M. Navarro (2014). "Did liberalising bar hours decrease traffic accidents". *Journal of Health Economics*, 35: 189–198.

Hafsteinsdottir, E. and L. Siciliani (2010). "DRG prospective payment systems: refine or not refine?" *Health Economics*, 19: 1226–1239.

Hahn, J. and R. Shi (2017). "Synthetic Control and Inference". *Econometrics*, 5:

Hernæs, Ø. (2018). "Activation against absenteeism - Evidence from a sickness insurance reform in Norway". *Journal of Health Economics*, 62: 60–68.

Hodgkin, D. and T. McGuire (1994). "Payment levels and hospital response to prospective payment". *Journal of Health Economics*, 13: 1–29.

Hu, L., R. Kaestner, B. Mazumder, S. Miller, and A. Wong (2018). "The effect of the affordable care act Medicaid expansions on financial wellbeing". *Journal of Public Economics*, 163: 99–112.

Imbens, G. W. and J. M. Wooldridge (2009). "Recent Developments in the Econometrics of Program Evaluation". *Journal of Economic Literature*, 47: 5–86.

Januleviciute, J., J. E. Askildsen, O. Kaarboe, L. Siciliani, and M. Sutton (2016). "How do Hospitals Respond to Price Changes? Evidence from Norway". *Health Economics*, 25: 620–636.

Kreif, N., R. Grieve, D. Hangartner, A. Turner, S. Nikolova, and M. Sutton (2016). "Examination of the synthetic control method for evaluating health policies with multiple treated units". *Health Economics*, 25: 1514–1528.

Lave, J. and R. Frank (1990). "Effect of the structure of hospital payment on length of stay". *Health Services Research*, 25: 327–347.

Long, M., J. Chesney, R. Ament, S. DesHarnais, S. Fleming, E. Kobrinski, and B. Marshall (1987). "The Effect of PPS on Hospital Product and Productivity". *Medical Care*, 25: 528–538.

Ma, C.-t. A. (1994). "Health Care Payment Systems: Cost and Quality incentives". *Journal of Economics & Management Strategy*, 3: 93–112.

Marini, G., M. Miraldo, R. Jacobs, and M. Goddard (2008). "Giving greater financial independence to hospitals - does it make a difference? The case of English NHS trusts". *Health Economics*, 17: 751–775.

Marla, S. and S. Stallard (2009). "Systematic review of day surgery for breast cancer". *International Journal of Surgery*, 7: 318–323.

McLennan, D., H. Barnes, M. Noble, J. Davis, E. Garrett, and C. Dibben (2011). *The English Indices of Deprivation 2010*. Department for Communities and Local Government.

Milstein, R. and J. Schreyögg (2016). "Pay for performance in the inpatient sector: A review of 34 P4P programs in 14 OECD countries". *Health Policy*, 120: 1125–1140.

Monitor & NHS England (2014). *2014/15 National Tariff Payment System*.

Moreno-Serra, R. and A. Wagstaff (2010). "System-wide impacts of hospital payment reforms: Evidence from Central and Eastern Europe and Central Asia". *Journal of Health Economics*, 29: 585–602.

NICE (2014). *Gallstone disease: Diagnosis and management of cholithiasis cholecystisis and choledocholithiasis: Clinical Guideline 188 Methods, evidence and recommendations*. Tech. rep. 88. National Institute for Health and Care Excellence.

O'Reilly, J., R. Busse, U. Häkkinen, Z. Or, A. Street, and M. Wiley (2012). "Paying for hospital care: the experience with implementing activity-based funding in five European countries". *Health Economics, Policy and Law*, 7: 73–101.

Olsen, C. B. and H. O. Melberg (2018). "Did adolescents in Norway respond to the elimination of copayments for general practitioner services?" *Health Economics*, 27: 1120–1130.

Paris, V., M. Devaux, and L. Wei (2010). *Health Systems Institutional Characteristics: A Survey of 29 OECD Countries*. OECD Health Working Paper 50. Organisation for Economic Cooperation and Development, Paris.

Rosko, M. and R. Broyles (1986). "The Impact of the New Jersey All-Payer DRG System". *Inquiry*, 23: 67–75.

Rothwell, P. M. (2005). "External validity of randomised controlled trials: 'To whom do the results of this trial apply?'" *The Lancet*, 365: 82–93.

Royal College of Surgeons of England (1985). *Commission on the Provision of Surgical Services. Guidelines for Day-Case Surgery*.

Salkever, D., D. Steinwachs, and A. Rupp (1986). "Hospital Cost and Efficiency under Per Service and Per Case Payment in Maryland: A Tale of the Carrot and the Stick". *Inquiry*, 23: 56–67.

Shin, E. (2019). "Hospital responses to price shocks under the prospective payment system". *Health Economics*, 28: 245–260.

Shleifer, A. (1985). "A Theory of Yardstick Competition". *RAND Journal of Economics*, 16: 319–327.

Shmueli, A., O. Intrator, and A. Israeli (2002). "The effects of introducing prospective payments to general hospitals on length of stay, quality of care, and hospitals' income: the early experience of Israel". *Social Science & Medicine*, 55: 981–989.

Siciliani, L. (2006). "Selection of treatment under prospective payment systems in the hospital sector". *Journal of Health Economics*, 25: 479–499.

Street, A. and A. Maynard (2007). "Activity based financing in England: the need for continual refinement of payments by results". *Health Economics, Policy and Law*, 2: 419–427.

Thompson, C. R. and M. McKee (2011). "An analysis of hospital capital planning and financing in three European countries: Using the principal-agent approach to identify the potential for economic problems". *Health Policy*, 99: 158–166.

Vaughan, J., K. Gurusamy, and B. R. Davidson (2013). "Day-surgery versus overnight stay surgery for laparoscopic cholecystectomy". *Cochrane Database of Systematic Reviews*, 7:

Verzulli, R., G. Fiorentini, M. L. Bruni, and C. Ugolini (2017). "Price Changes in Regulated Healthcare Markets: Do Public Hospitals Respond and How?" *Health Economics*, 26: 1429–1446.

7 Appendix

Table A1: Means of patient characteristics

#	BPT	Age	Male	Deprivation score	Elixhauser score	Past emergency admission
1	Cholecystectomy	49.9	0.22	0.16	0.97	0.43
2	Simple mastectomy	50.9	0.17	0.13	0.54	0.09
3	Sentinel node mapping and resection	59.0	0.10	0.13	0.99	0.08
4	Operations to manage female incontinence	53.3	0.00	0.14	0.73	0.07
5	Endoscopic prostate resection	72.1	1.00	0.13	1.78	0.38
6	Laser prostate resection	71.4	1.00	0.13	1.56	0.37
7	Hernia repair	58.3	0.85	0.14	0.86	0.11
8	Therapeutic arthroscopy of shoulder	56.1	0.50	0.14	0.95	0.07
9	Bunion operation	56.4	0.16	0.14	0.72	0.05
10	Fasciectomy	64.6	0.78	0.13	0.81	0.06
11	Tonsillectomy	32.0	0.37	0.16	0.37	0.17
12	Septoplasty	41.2	0.69	0.15	0.42	0.06
13	Tympanoplasty	42.4	0.50	0.16	0.15	0.06
14	Epileptic seizure	53.5	0.54	0.18	3.57	0.59
15	Acute headache	45.9	0.35	0.17	1.22	0.30
16	Asthma	47.1	0.30	0.19	2.55	0.40
17	Lower respiratory tract infections without COPD	51.7	0.44	0.17	0.70	0.26
18	Pulmonary embolism	62.3	0.47	0.14	3.03	0.36
19	Chest pain	59.3	0.53	0.17	2.22	0.37
20	Appendicular fractures not requiring fixation	63.4	0.41	0.16	1.61	0.26
21	Cellulitis	57.0	0.56	0.16	1.66	0.31
22	Renal/ureteric stones	45.8	0.69	0.17	0.74	0.27
23	Deep vein thrombosis	61.8	0.50	0.16	2.03	0.43
24	Deliberate self-harm	39.1	0.43	0.20	2.19	0.44
25	Falls including syncope or collapse	67.6	0.52	0.16	2.46	0.37
26	Community acquired pneumonia	51.8	0.50	0.16	0.63	0.22
27	Arrhythmia	68.1	0.48	0.14	3.42	0.39
28	Minor head injury	54.9	0.56	0.18	1.63	0.33
29	Low risk pubic rami	81.3	0.15	0.14	2.43	0.37
30	Bladder outflow obstruction	68.5	0.81	0.15	2.15	0.39
31	Anemia	69.7	0.36	0.17	3.94	0.38
32	Abdominal pain	47.7	0.35	0.17	1.51	0.39

Notes: See Section 3 for variable definitions.

Table A2: Volume of incentivised activity and % growth over time

#	BPT	Volume of activity									Average growth per annum		
		2006	2007	2008	2009	2010	2011	2012	2013	2014	Pre-policy	Post-policy	Total
1	Cholecystectomy	9,751	9,997	10,253	12,087	12,244	12,842	12,327	13,064	12,914	4.8%	1.1%	4.1%
2	Simple mastectomy	4,417	4,393	4,437	4,430	3,949	3,713	3,667	3,821	3,801	-1.8%	0.6%	-1.7%
3	Sentinel node mapping and resection	4,982	6,048	9,513	11,842	15,190	17,224	19,504	21,408	23,131	34.1%	8.6%	45.5%
4	Operations to manage female incontinence	8,623	13,751	14,138	13,803	13,380	12,891	11,935	11,853	9,586	9.2%	-6.4%	1.4%
5	Endoscopic prostate resection	6,654	6,288	5,856	6,312	6,111	6,146	6,102	5,934	5,458	-1.4%	-2.8%	-2.2%
6	Laser prostate resection	15,563	17,051	17,381	16,453	15,531	15,453	15,032	15,505	14,867	0.0%	-0.9%	-0.6%
7	Hernia repair	89,900	94,914	92,737	89,731	90,208	94,571	92,502	97,968	98,148	0.1%	0.9%	1.1%
8	Therapeutic arthroscopy of shoulder	3,542	2,780	1,572	22,223	29,176	33,607	33,411	35,526	36,886	120.6%	2.4%	117.7%
9	Bunion operation	10,741	12,882	13,985	14,757	16,811	16,848	14,753	14,850	14,771	9.4%	-3.1%	4.7%
10	Fasciectomy	11,813	10,551	9,631	9,128	9,174	8,865	8,526	8,360	7,950	-3.7%	-2.6%	-4.1%
11	Tonsillectomy	16,456	16,693	16,123	16,148	15,301	15,138	15,830	17,066	17,000	-1.1%	2.5%	0.4%
12	Septoplasty	19,158	19,511	19,375	19,039	19,542	19,391	18,580	19,527	19,078	0.2%	0.9%	-0.1%
13	Tympanoplasty	9,624	10,284	9,728	9,428	8,910	7,677	7,204	7,104	6,899	-3.1%	-1.4%	-3.5%
14	Epileptic seizure	41,716	42,427	45,337	47,181	35,170	47,479	47,477	47,477	46,671	2.0%	-0.6%	1.5%
15	Acute headache	40,674	43,194	49,439	54,866	55,835	56,501	58,532	62,290	63,113	5.6%	2.6%	6.9%
16	Asthma	32,030	30,236	33,114	30,523	30,132	26,555	29,690	27,871	31,879	-2.4%	2.5%	-0.1%
17	Lower respiratory tract infections without COPD	15,168	14,128	14,023	10,411	10,235	8,867	10,281	8,689	9,873	-5.9%	-1.3%	-4.4%
18	Pulmonary embolism	9,170	10,033	10,849	11,689	11,014	11,394	12,638	12,801	12,826	3.5%	0.5%	5.0%
19	Chest pain	248,882	243,410	258,997	264,983	198,080	259,147	253,091	254,538	243,264	0.6%	-1.3%	-0.3%
20	Appendicular fractures not requiring fixation	35,950	38,678	40,348	43,422	40,252	38,783	37,659	38,221	38,857	1.1%	1.1%	1.0%
21	Cellulitis	33,305	32,478	32,906	32,893	24,675	31,633	30,479	31,713	33,229	-0.7%	3.0%	0.0%
22	Renal/ureteric stones	26,553	25,805	26,889	29,182	28,817	27,891	26,667	27,627	28,137	0.7%	1.8%	0.7%
23	Deep vein thrombosis	20,314	20,763	22,313	22,233	19,842	17,060	16,686	17,135	17,770	-2.3%	2.2%	-1.6%
24	Deliberate self-harm	85,936	88,754	91,402	93,432	96,790	97,304	91,016	94,837	88,189	1.9%	-1.0%	0.3%
25	Falls including syncope or collapse	61,251	60,699	66,399	66,905	65,019	60,617	55,991	54,337	51,485	-0.1%	-2.7%	-2.0%
26	Community acquired pneumonia	13,717	13,161	13,160	11,998	12,514	10,483	11,326	9,377	10,914	-2.2%	8.2%	-2.6%
27	Arrhythmia	87,039	89,842	91,941	97,052	93,371	94,086	95,232	97,292	97,223	1.2%	0.0%	1.5%
28	Minor head injury	21,092	19,196	18,336	18,700	15,914	14,914	13,003	13,115	12,416	-4.8%	-2.7%	-5.1%
29	Low risk pubic rami	5,374	5,799	5,945	6,521	6,414	6,712	7,230	7,645	7,853	4.3%	1.4%	5.8%
30	Bladder outflow obstruction	13,584	13,610	13,567	13,472	11,898	11,529	10,446	9,467	8,674	-2.9%	-4.2%	-4.5%
31	Anaemia	9,387	10,839	11,731	13,100	11,435	12,241	13,088	13,711	14,189	4.9%	1.7%	6.4%
32	Abdominal pain	174,494	173,899	185,860	197,229	199,249	197,419	196,163	199,559	198,755	1.6%	-0.2%	1.7%