

This is a repository copy of *Paying for efficiency: Incentivising same-day discharges in the English NHS*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/135067/>

Version: Published Version

Monograph:

Gaughan, James Michael orcid.org/0000-0002-8409-140X, Gutacker, Nils orcid.org/0000-0002-2833-0621, Grasic, Katja et al. (3 more authors) (2018) *Paying for efficiency: Incentivising same-day discharges in the English NHS*. Report. CHE Research Paper . Centre for Health Economics, University of York , York.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

ESHCRU

Economics of
Social and Health Care
Research Unit



Centre For Health Economics



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

James Gaughan, Nils Gutacker,
Katja Grasic, Noemi Kreif, Luigi Siciliani,
Andrew Street

CHE Research Paper 157

Paying for Efficiency: Incentivising same-day discharges in the English NHS

^aJames Gaughan

^aNils Gutacker

^aKatja Grasic

^aNoemi Kreif

^bLuigi Siciliani

^cAndrew Street

^aCentre for Health Economics, University of York, UK

^bDepartment of Economics and Related Studies, University of York, York, UK

^cDepartment of Health Policy, The London School of Economics and Political Science, UK

August 2018

Background to series

CHE Discussion Papers (DPs) began publication in 1983 as a means of making current research material more widely available to health economists and other potential users. So as to speed up the dissemination process, papers were originally published by CHE and distributed by post to a worldwide readership.

The CHE Research Paper series takes over that function and provides access to current research output via web-based publication, although hard copy will continue to be available (but subject to charge).

Acknowledgements

We are grateful for comments and suggestions from Hugh McLeod, Kian Chin, Paul Twigg, as well as those received during presentations at the 2016 Winter Health Economists' Study Group meeting, the 3rd German Health Econometrics Workshop, the 2017 HEB-HERO Health Economics Workshop and during seminars at the RWI-Leibniz Institute, Germany and the Universities of Lancaster, Hamburg and Duisburg-Essen. This work reported here is based on independent research commissioned and funded by the 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, arm's length bodies or other government departments. The Hospital Episode Statistics are copyright © 2006-2014, re-used with the permission of NHS Digital. All rights reserved.

No ethical approval was needed.

Further copies

Only the latest electronic copy of our reports should be cited. Copies of this paper are freely available to download from the CHE website www.york.ac.uk/che/publications/. Access to downloaded material is provided on the understanding that it is intended for personal use. Copies of downloaded papers may be distributed to third parties subject to the proviso that the CHE publication source is properly acknowledged and that such distribution is not subject to any payment.

Printed copies are available on request at a charge of £5.00 per copy. Please contact the CHE Publications Office, email che-pub@york.ac.uk, telephone 01904 321405 for further details.

Centre for Health Economics
Alcuin College
University of York
York,
YO10 5DD, UK
www.york.ac.uk/che

©James Gaughan, Nils Gutacker, Katja Grasic, Noemi Kreif, Luigi Siciliani, Andrew Street

Abstract

We study a pay-for-efficiency scheme that encourages hospitals to admit and discharge patients on the same calendar day where clinically appropriate. Since 2010, hospitals in the English NHS receive 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 for 191 conditions for which same-day discharge is clinically appropriate — of which 32 are incentivised — during 2006-2014. Using interrupted time series, differences-in-differences and synthetic control methods, we find that the policy generally had a positive effect on planned conditions with a statistically significant effect in about a third of conditions. The results are more mixed for emergency conditions. The median elasticity (across all 32 conditions) is 0.09 but above one for six conditions. Condition-specific design features explain some, but not all, of the differential responses.

JEL: D22, I11

Keywords: Pay for Performance; Best Practice Tariff; day surgery; same-day discharge; policy evaluation

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) in the US or a Healthcare Resource Group (HRG) in England, 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, this being an important cost driver. For some patients it may be possible to reduce length of stay to zero, specifically those for whom care can be provided safely² within an ambulatory 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, it might be to the patient's benefit. The British Association of Day Surgery (BADS) has recommended SDD for nearly 180 types of planned surgery (BADS 2006) and the British Association for Ambulatory Emergency Care (BAAEC) has identified a range of conditions that require urgent care but where a subsequent overnight stay for observation is generally considered unnecessary (BAAEC 2014). Implementing these recommendations makes financial sense in the English National Health Service (NHS): for patients allocated to the same HRG, hospitals are paid the same amount for SDD treatment as for treating those who have an overnight hospital stay, despite the cost of providing SDD care being substantially lower (Street and Maynard 2007).³ This should give hospitals a financial incentive to treat patients on an SDD basis whenever clinically appropriate.

Despite these recommendations and financial incentives, SDD rates are lower than is clinically recommended for a wide range of treatments (Department of Health 2010)(see also Figure 1). The reasons for low rates may relate to reluctance by doctors or to features of the hospital that constrain the ability to offer care on an SDD basis. One way to encourage doctors and hospitals to address these reasons is by increasing the SDD price, and this has been the approach taken in England. A payment reform known as the SDD *Best Practice Tariff* (BPT) involves paying a higher price for SDD than for care that involves an overnight or longer stay in hospital and has been applied to 32 different conditions.⁴ The SDD payment policy is unusual in that it pays more for the less costly treatment, making it distinct from the usual form of PPS in which prices are set at average cost (Shleifer 1985).

We investigate whether hospitals responded to the SDD incentive scheme and, in so doing, we contribute to two related strands of literature. First, we contribute to studies that focus on the effect of price changes on treatment choices. These find that physicians are willing to change their care patterns in response to financial incentives (see Chandra et al. (2011) for a recent review of this literature).

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 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>)

⁴ Formally, planned and emergency SDD care is incentivised through two different BPTs. However, the design of both BPTs is identical and we therefore refer to both as one BPT.

For example, a growing body of literature has shown that obstetricians respond to changes in the profitability of caesarean section compared to vaginal birth by amending their treatment thresholds for the invasive surgical procedure (e.g. Gruber et al. 1999; Allin et al. 2015; Foo et al. 2017). For planned hip replacement, Papanicolas and McGuire (2015) found that more generous reimbursement for un-cemented relative to cemented implants in the English NHS led to greater provision of the former, despite a clinical recommendation in favour of the latter. Finally, Farrar et al. (2009) evaluated the introduction of PPS in England and found that it led to 0.4-0.8% more planned surgery being performed as SDD as well as an overall reduction in length of stay.

Our work also contributes to a second strand of literature evaluating pay-for-performance (P4P) programmes. A review of 34 hospital sector P4P schemes in the US and other OECD countries finds the effects to be generally modest in size, short lived and sometimes associated with unintended consequences (Milstein and Schreyögg 2016). The authors argue that the effectiveness of a P4P scheme is associated with the size of the incentive and that they are most appropriate for emergency care, where hospitals have less opportunity to select patients. Most P4P schemes focus on incentivising quality, either through rewarding health outcomes or process measures of quality. But the P4P policy we examine is distinct in that it incentivises efficiency so may be better termed a pay-for-*efficiency* (P4E) programme.

We offer several novelties to the existing literature. First, we analyse an unusual payment policy in which English hospitals are paid a bonus BPT for treating patients on an SDD basis. This policy explicitly and intentionally overpays hospitals for the cheapest care pathway, the objective being to stimulate take-up and improve efficiency. Our study extends a previous study by Allen et al. (2016) which evaluates the short-term effects of this P4E policy for cholecystectomy patients in England. That study used a DiD approach with a control group of all non-incentivised procedures recommended for SDD and found an increase in SDD rates of 5.8 percentage points (pp) in the first 12 months following the policy introduction. We extend that study in two ways. Firstly, instead of just one condition, we examine 32 conditions to which a similar bonus policy applied. This allows us to assess the generalisability of the policy by, in effect, conducting 32 separate experiments. Secondly, we examine longer-term effects, up to five years after the introduction of the bonus payment policy, which allows us to examine temporal responses.

Second, a distinctive feature is that the SDD incentive scheme was high-powered. The size of the bonus was economically significant, varying from 8% to 66% more than for an overnight hospital stay. This price differential compounds the cost advantage, which varied from 23% to 71% lower for SDD than for an overnight hospital stay in the pre-policy period. These incentives are much larger than those associated with most other P4P schemes, which are often around 5% (Cashin et al. 2014). The analysis can therefore shed light on whether limited responsiveness to P4P schemes as documented in literature is simply due to the small size of the bonus.

Third, we apply and compare three different econometric strategies, namely interrupted time series (ITS) analysis, difference-in-difference (DiD) methods, and synthetic control (SC) methods pioneered by Abadie and Gardeazabal (2003) and Abadie et al. (2010). While DiD methods are commonly applied in health policy evaluations, SC methods are a fairly recent addition to our analytical armoury but are receiving increasing attention in the wider economic literature (e.g. Billmeier and Nannicini 2013; Bharadwaj et al. 2014; Green et al. 2014; Kreif et al. 2016; Acemoglu et al. 2017). Sometimes it is not possible to apply DiD or SC methods because of the need to identify appropriate control groups. In this study, because we examine the same type of policy applied to 32 different conditions, we have subsets of conditions to which either all three or just a subset of the methods can be applied. Consequently we are able to compare

results from different methods for subsets of conditions we analyse, according to which underpinning methodological assumptions are satisfied for each condition. This serves as a robustness check for our findings.

Our results can be summarised as follows: Reassuringly, we find similar results to Allen et al. (2016) for cholecystectomy but, disappointingly from a policy perspective, it turns out that the bonus has the largest effect for this condition and its impact cannot be generalised. The BPT policy led to a statistically significant increase in SDD rates of 4-10pp for four out of 13 planned conditions. Results for emergency conditions are more mixed with four positive and three negative statistically significant effects. Furthermore, the magnitudes of effects for emergency conditions are generally smaller, ranging from +6pp to -6pp where statistically significant. 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. We find no clear temporal pattern of policy response across conditions, again making it difficult to draw general policy conclusions. Findings are broadly robust to the use of different analytical approaches.

The paper 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 theoretical predictions

The English NHS is funded by general taxation and patients face no charges for hospital care. Residents have to be registered with a general practitioner, who act as gatekeepers and can refer patients for planned inpatient care to any licenced hospital in England. Patients can be admitted for emergency care via a hospital's Accident & Emergency department or by direct referral from their general practitioner. Most hospitals are publicly owned, although a small number of private hospitals also 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 tariff for treating NHS-funded patients, differentiated by 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).⁵ Initially limited to a small number of planned conditions, PPS has been extended progressively over time and now covers most hospital activity.

We start by describing the construction of prices in the pre-policy period prior to the introduction of the SDD policy. We denote the pre-policy period with $\alpha = 0$ and the post-policy period as $\alpha = 1$. While the SDD policy was introduced for different patient groups at different times, we analyse each group individually.

The tariff 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, $\bar{C}_{k-3,g}$.⁶ More formally, $\bar{C}_{k-3,g} = (\sum_{j=1}^J (C_{k-3,j,g} \times N_{k-3,j,g}) / \sum_{j=1}^J N_{k-3,j,g})$, where $j = 1, \dots, J$ denotes the hospital, $N_{k-3,j,g}$ is the number of patients for a given hospital j , and $C_{k-3,g}$ is the average cost of patients in hospital j . Reimbursement is further adjusted to account for inflation (I) and expected efficiency improvement (E) factors.⁷ Therefore, the pre-policy price $P_{0,k,g} = \bar{C}_{k-3,g} \times I_k \times E_k$, with $I_k > 1$ and $E_k < 1$.

For most planned treatment, patients admitted and discharged on the same day (*SDD*) attract the same payment as overnight stays (*ON*). Therefore, $P_{0,k,g} = P_{0,k,g}^{SDD} = P_{0,k,g}^{ON}$ 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 $0 < \lambda \leq 1$ 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. The short-stay adjustment is aimed at reducing the incentive to admit less severe patients for observation rather than intervention. Therefore, emergency care including at least one overnight stay has a price constructed equivalently to planned care $P_{0,k,g}^{ON} = P_{0,k,g}$ while $P_{0,k,g}^{SDD} = \lambda P_{0,k,g}$.

The BADS and BAAEC both produce directories listing 191 clinical conditions (i.e. specific diagnoses or surgical treatments) that are deemed suitable for SDD and a recommended rate (RR) of SDD that is considered safe and appropriate (BADS 2006; BAAEC 2014). The directories represent a

⁵ The policy was originally known as 'Payment by Results' and has since been renamed as 'National Tariff Payment System'.

⁶ 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.

⁷ 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.

clinical consensus about the appropriate level of SDD. From 2010, the English Department of Health has gradually introduced explicit financial incentives (SDD BPTs) for specific conditions from these directories.⁸ These incentives apply to all providers of NHS-funded care. The selection and design of SDDs was informed by discussions with clinical stakeholders and varies across clinical areas (Department of Health 2007). New conditions to be incentivised are announced six months in advance of introduction. The general criteria for potential selection are volume ($>5,000$ patients/year)⁹, the national SDD rate being below the RR for this condition, and evidence of variation in the SDD rate across hospitals (Department of Health 2010). Not all clinical conditions meeting these general criteria have an SDD incentive but by April 2014, 13 planned and 19 emergency conditions were covered by SDD incentives (Monitor and NHS England 2014).

A condition incentivised by an SDD BPT has two prices such that $P_{1,k,g}^{SDD} > P_{1,k,g}^{ON}$ for both planned and emergency care. For example, under this scheme in 2010, hospitals are paid approximately £329 (or 24%) more for a same day discharge than an overnight stay for planned cholecystectomy (gall bladder removal)(Department of Health 2010). This structure is common to all 32 SDD BPTs. However, the absolute and relative size of the differential varies considerably and range from 8% to 66% of the overnight admission price. After their introduction bonuses were approximately stable over time.¹⁰ For planned care, a higher price is only paid if the patient was scheduled to be treated as a day case in advance of admission. Therefore, the price for a patient discharged on the same day but not admitted as a day case is the same as an overnight stay.

Table 1 provides an overview of the incentivised SDD conditions, the financial year in which the incentive was introduced¹¹, the hospital reimbursement with and without the SDD incentive, the average cost of care reported by NHS hospitals in the year prior to the policy, and in addition the SDD rate and the number of patients eligible in the twelve months prior to announcement of the incentive for that group.

2.1. Hospital incentives

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

The aim of the SDD pricing policy was to increase the rate of SDD towards the recommended rate by introducing a financial incentive for hospitals. We illustrate this incentive for *planned* day case surgery first. The profit function, denoted with π in the pre-policy and the post-policy period is given respectively

⁸ In some cases, additional exclusion criteria are applied to limit the scope of the SDD BPT to non-complex patients. In these cases, the group of patients with incentivised tariffs attached is a subset of those given in relevant directories and recommended rates can be considered a lower bound of what is clinically appropriate.

⁹ One noteworthy exception is 'simple mastectomy' which has been incentivised since 2011 despite an annual volume of about 4,000 patients.

¹⁰ 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 BPT 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.

¹¹ 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 | 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 | Shoulder decompression | 2011 | 80 | 26,836 | 49 | 2,172 | 2,172 | 0 | 2,253 | 2,053 | 200 | 1,319 | 2,047 | -729 |
| 9 | Bunion operation | 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 | Epilepsy | 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 | Respiratory | 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 | 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 | 2012 | 90 | 62,230 | 32 | 443 | 985 | -542 | 751 | 546 | 205 | 401 | 994 | -593 |
| 26 | Pneumonia | 2013 | 30 | 11,121 | 19 | 609 | 1,353 | -744 | 1,136 | 936 | 200 | 447 | 1,374 | -927 |
| 27 | Fibrillation | 2013 | 60 | 96,203 | 26 | 682 | 1,588 | -906 | 1,242 | 1,026 | 216 | 465 | 1,373 | -908 |
| 28 | Head injury | 2013 | 60 | 13,976 | 53 | 477 | 546 | -69 | 698 | 453 | 245 | 424 | 1,074 | -649 |
| 29 | Pelvis fracture | 2013 | 90 | 6,935 | 8 | 344 | 1,374 | -1,030 | 1,711 | 1,466 | 245 | 971 | 3,861 | -2,890 |
| 30 | Bladder outflow | 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.

by

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

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

The difference in profit before and after the policy is:

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

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 tariff. The third term is positive if the SDD price induces an increase in the SDD rate, which are 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; Hafsteinsdottir and Siciliani 2010, for more formal theoretical models). 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 day case treatments, but the overall effect on profits also depends on the reduction in the base tariff. A similar analysis holds for emergency care where the only difference is that pre-policy the tariff was higher for overnight treatments, i.e. $(P_1^{SDD} - P_1^{ON}) > P_0^{SDD} - P_0^{ON} < 0$.

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

As regards motivation, slow uptake of SDD may reflect poor dissemination about best practice. Doctors may not be aware of or may doubt the evidence that SDD is as safe as traditional practice involving overnight admission. 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.

Moreover, the hospital in which the doctor works may be constrained in its ability to extend SDD to more patients. While many SDD treatments can be performed in a normal hospital setting, making SDD standard practice may require building new facilities or repurposing existing hospital units that are

devoted to SDD care. If so, expanding the volume of SDDs may require a long-term capital investment. Some English hospitals may not undertake this investment, particularly those that face greater borrowing constraints that restrict their access to capital funds (Marini et al. 2008; Thompson and McKee 2011).

2.2. Welfare

We conclude this section by discussing welfare implications. We discuss welfare under two perspectives. First, we define welfare as the difference in patient benefits minus provider costs. It has been argued that an increase in SDD rate will not harm patients as long as it remains below the RR. Under this assumption, the introduction of the SDD price incentive will have no effect on patient benefits if SDD rates increase, so that the effect of welfare is driven by its effect on costs. As argued above, an increase in SDD will reduce costs under minimal regularity conditions. We can therefore conclude that the SDD pricing policy is welfare improving, and that the size of the welfare gain increases with the number of SDDs (up to the RR).

Second, we take the purchaser perspective, and define welfare more narrowly as the difference between patient benefit and the transfer to the provider. Since patient benefit does not differ between SDD and ON, the effect of the SDD price on purchaser welfare is, as shown above, given by its effect on the overall transfer, and equal to $(P_1^{SDD} - P_1^{ON})N_1^{SDD} - N(P_0 - P_1^{ON})$. This suggests that the purchaser is always better off when the SDD price is introduced as long as it sufficiently reduces the ON tariff to compensate for the increase in the transfer to the provider due to the increase in SDD price.

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 clinical and socio-demographic characteristics, the admission pathway and its timings, and whether care was scheduled as SDD in advance (planned admissions only). The outcome of interest is constructed as a binary variable that takes the value of 1 if the patient is admitted and discharged on the same calendar day, and zero otherwise.

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 the observed rate of SDD and the gap between SDD rate and RR, i.e. the potential for growth.

In our empirical analyses we control for potential changes in patient complexity over time that may explain observed changes in SDD rates. We construct a set of risk-adjustment variables from the HES dataset including 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 (McLennan et al. 2011) for the patients' lower layer super output area of residence.

Hospitals are consulted on any changes to the payment system — including the introduction of new BPTs — approximately six months prior to the change. This gives them time and opportunity to adapt to the new policy before the actual implementation, which may bias observed pre-policy outcomes. We therefore exclude data for all patients treated in the six months prior to the condition being incentivised. Also, for some conditions eligibility criteria were refined over time to restrict the incentive to a more tightly defined patient population. In these instances, we apply the criteria that were valid when the financial incentive first applied to this condition 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 condition are available in Table A1 in the Appendix.

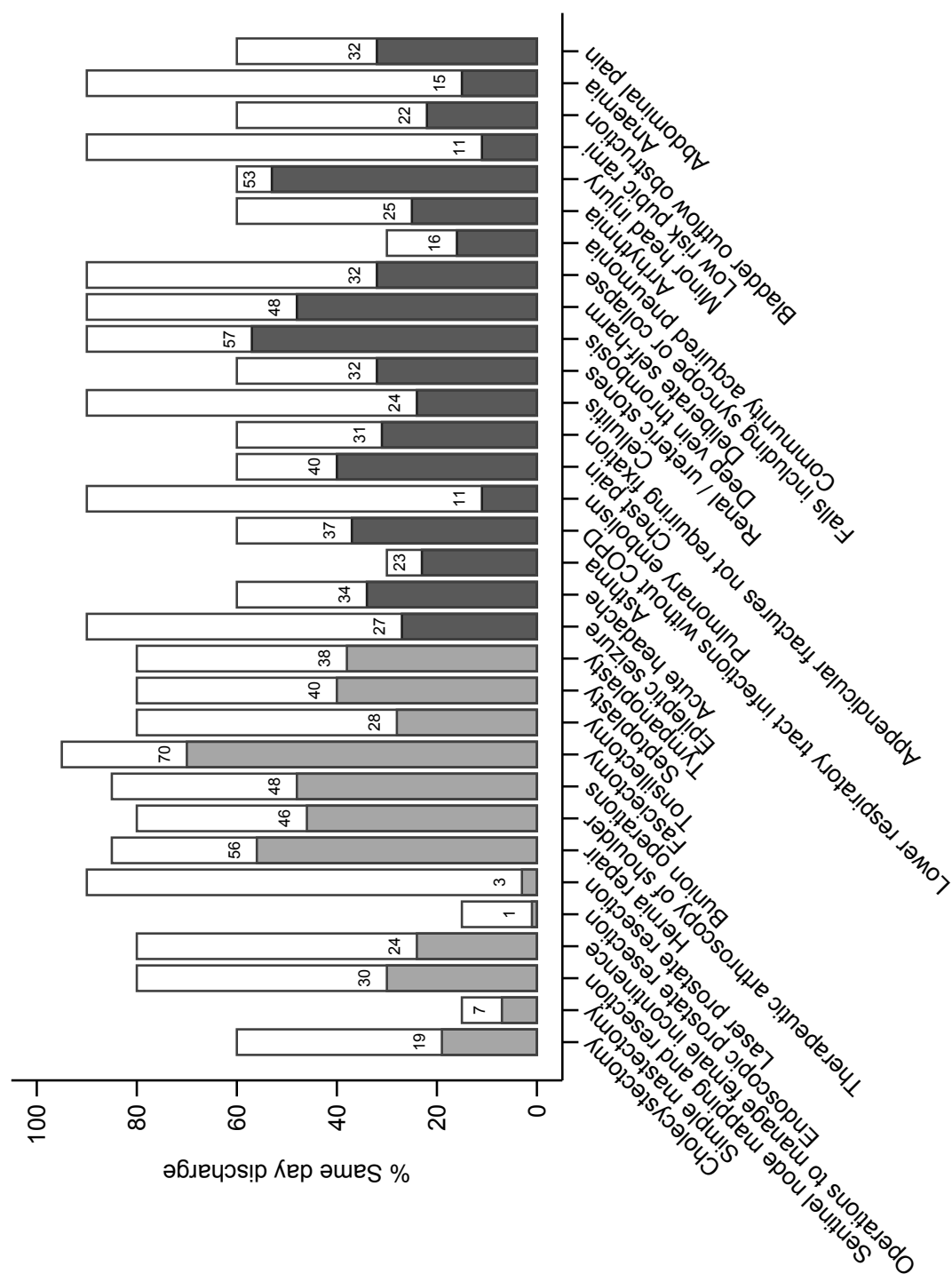


Figure 1: Observed SDD rate (grey bar) and recommended rate (white bar) for incentivised conditions in 2009

4. Methods

Our empirical analysis seeks to estimate the impact of the SDD pricing policy on the probability of a patient being discharged on the same day as admission.¹² Separate models are estimated for each of the 32 incentivised conditions.

We follow the potential outcome framework developed by Rubin and commonly applied in the policy evaluation literature (Rubin 1974; Abadie and Imbens 2006). For every patient, we define two potential outcomes: Y_{it}^1 is the outcome that a patient i would realise in month t if the SDD incentive was in effect (potential outcome under treatment) and Y_{it}^0 is the potential outcome that the same patient would realise without the SDD incentive (potential outcome under control). For patients who received care for one of the 32 incentivised conditions, their observed outcomes before the introduction of the SDD policy ($t < t^{BPT}$) correspond to their potential outcome under control, where t^{BPT} is defined as the month when the SDD BPT was introduced. After the introduction of the SDD BPT, their observed outcomes correspond to their potential outcomes under treatment. By contrast, for patients who received care for non-incentivised conditions (used as control groups in Sections 4.2 and 4.3), observed outcomes correspond to potential outcomes under control throughout the entire study period.

Policy interest is in the average treatment effect on the treated (ATT), i.e. the patients who were treated as an SDD for incentivised conditions after the introduction of the policy, defined as $E[Y_{it}^1 - Y_{it}^0 | t \geq t^{BPT}]$. While Y_{it}^1 is observed for these patients, the potential counterfactual outcome under control Y_{it}^0 is unobserved. We employ three different analytical approaches to estimate the expected counterfactual outcome $E[Y_{it}^0]$.

4.1. Interrupted time series analysis

Our first analytical approach employs *interrupted time series* (ITS) analysis. The identifying assumption of the ITS design is that a linear pre-policy trend in the proportion of SDD would have continued uninterrupted in the absence of the SDD BPT policy. Therefore, the trend in the observed values of Y_{it} for $t < t^{BPT}$ can be used to construct the counterfactual outcome Y_{it}^0 for $t \geq t^{BPT}$.

ITS analysis uses segmented regression techniques to test for structural breaks in the linear time trend when the SDD policy is introduced. The ITS specification commonly used in empirical policy evaluations allows for a single break, which may manifest as an immediate shift in the proportion of SDD and/or a homogeneous change in its trend. We extend this base specification to allow for heterogeneous effects in each of the $k = 1, \dots, K$ post-policy years following the introduction of the SDD policy and specify the

¹² 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 BPT (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.

regression model as

$$Pr[Y_{ijt} = 1] = \alpha_0 + \alpha_1 M_t + \sum_{k=1}^K [\gamma_k D_k + \delta_k (D_k \times M_t)] + \sum_{s=1}^4 \nu_s Q_s + \sum_{j=1}^J \theta_j H_j + (X_i \times Z_t)' \xi \quad (4)$$

where Y_{ijt} is a dummy variable taking the value of 1 if patient $i = 1, \dots, I$ treated in hospital $j = 1, \dots, J$ in month $t = 1, \dots, T$ was admitted and discharged on the same day and the value of 0 if the patient was admitted and stayed at least one night in hospital. The variable M_t is a continuous measure of time in months.

D_k are dummy variables which take the value of 1 in each of the $k = 1, \dots, K$ post-policy years and zero otherwise. The coefficients γ_k and δ_k measure shifts and changes in trend in the proportion of SDD in each of the post-SDD years, respectively. Our model thus allows 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 increasing the proportion of patients treated on an SDD basis.

Q_s is a vector of seasonal (quarter) dummies, e.g. to allow for winter effects. H_j is a vector of hospital dummies, which capture unobserved time-invariant differences amongst hospitals (e.g. management quality, local demand) in the propensity to discharge patients the same day.

The adoption of SDD practice is likely to differ according to patient characteristics, with more severely ill patients less likely to be suitable for discharge on the same day that they receive treatment. Failure to account for patient case-mix may lead to biased estimates of the policy parameters if there are case-mix changes over time or if hospitals respond differently to the incentive for different patient groups. We address this concern by interacting a vector of patient characteristics X_i with $Z_t = [M_t, (D_k \times M_t)]$. As a result, trends in SDD rates can vary with patient severity and, therefore, the policy parameters can also vary across patient groups.

The ATT of the SDD BPT in year k for the baseline patient (when all elements of X_i equal zero), defined as τ_k , is calculated at the mid-point of each year k and given by

$$\tau_k = \gamma_k + \frac{1}{2} \delta_k \quad (5)$$

where γ_k denotes the level change in the SDD rate in the year k relative to the level implied by the pre-policy trend and δ_k is the change in its average monthly growth rate in the same years (relative to the counterfactual growth rate α_1). We calculate separate estimates of τ_k for each patient group defined by X_i and then average over the distribution of patients treated in each year k .

The key focus of this study is the ATT calculated over the entire post-policy period, which we define as $\bar{\tau}$, and is given by

$$\bar{\tau} = \frac{1}{N} \sum_{k=1}^K \tau_k N_k \quad (6)$$

where N_k is the number of patients in year k and $N = \sum_{k=1}^K N_k$.

All models are estimated as linear probability models with Huber-White robust standard errors of the model coefficients. The corresponding standard errors of policy parameters of interest are calculated using the delta method.

4.2. Difference-in-difference analysis

A key assumption of the ITS model is that the pre-policy trend is an unbiased estimate of the counterfactual Y_{it}^0 for the post-policy period. In other words, the trend in the proportion of SDD observed before the policy change would have continued afterwards if the intervention had not come into effect. This assumption may not hold if other concurrent events in the post-policy period affect the trend in SDD rates.

We relax this assumption by employing a *difference-in-difference* (DiD) strategy. We construct Y_{it}^0 based on the observed outcomes of a control group that is not affected by the SDD policy but is subject to the same external influences and would respond similarly to them. These requirements imply that both the intervention and the control show parallel trends in the average Y_{it} prior to the policy introduction. After accounting for differences between intervention and control groups in levels of expected outcomes prior to the policy introduction, any further difference in levels after the policy introduction can be interpreted as average effects of the SDD policy.

We estimate the following specification

$$\begin{aligned} Pr[Y_{ijt} = 1] = & \beta_0 + \beta_1 BPT_i + \sum_{k=1}^K [\gamma_k D_k + \mu_k (D_k \times BPT_i)] + \sum_{s=1}^4 [\nu_s Q_s + \varphi (Q_s \times BPT_i)] \\ & + \sum_{j=1}^J [\theta_j H_j + \omega_j (H_j \times BPT_i)] + (\mathbf{X}_i \times \mathbf{V}_i)' \boldsymbol{\xi} \end{aligned} \quad (7)$$

where BPT_i is a dummy variable that takes the value of 1 for patients in the intervention group and 0 for patients in the control group. All other variables are defined as in Section 4.1, except for V_i which is a matrix composed of BPT_i , D_k and $D_k * BPT_i$. This is analogous to $(\mathbf{X}_i \times \mathbf{Z}_t)$ in the ITS analysis and is, again, designed to capture changes in case-mix over time. We allow for hospital fixed effects to vary between the intervention and the control group to account for any differences in a hospital's relative propensity to discharge patients with different clinical conditions on the same day.¹³

The effect of the SDD policy in year k for the baseline patient is now given by $\tau_k = \mu_k$ and the calculation of the policy parameters proceeds as outlined before.

We select a separate control group for each incentivised condition. We consider as potential control groups all non-incentivised conditions from the clinical directories that follow the same admission pathway (planned or emergency), have a $RR \pm 15\%$ of the intervention group (see also Allen et al. 2016) and have at least, on average, 100 admissions per calendar month over the pre-policy period. Furthermore, to meet the assumptions of the DiD approach, we only consider control groups that show a similar trend in the proportion of SDDs prior to the introduction of the pricing policy, defined as $(\alpha_1^{BPT})/(\alpha_1^{Control}) = [0.9, 1.1]$

¹³ For example, a hospital may be 5pp more likely than the average hospital to discharge patients in the intervention group on the same day and 12pp more likely to do so for patients in the control group. In this case, forcing a common hospital fixed effect for both groups would be inappropriate.

with estimates of α_1 obtained from separate ITS regressions. Where multiple control groups meet these criteria, we use the control group with the most similar pre-policy level, i.e. $\min|\alpha_0^{BPT} - \alpha_0^{Control}|$.

4.3. Synthetic control analysis

DiD models are commonly applied using a single control group. In our study, we consider 32 incentivised conditions. For some of these there might be more than one potential control group that satisfies the selection criteria. For other incentivised conditions we might find no controls. We therefore also apply the *synthetic control* (SC) method. In short, the SC method allows evaluating the effect of a policy on a single treated unit (e.g. a country, region or, as in our case, an incentivised condition) by employing an algorithm to select a weighted combination of potential control units. Weights are chosen to minimise the difference from the intervention unit in terms of observed outcomes and predictors of outcome in the pre-policy period. Under a number of assumptions, including 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, the post-policy outcomes of the SC group can be interpreted as the counterfactual outcome of the intervention group. The difference between observed and counterfactual outcomes provides an estimate of the ATT.

As regards our study, there are two advantages of the SC method over the DiD method. First, SC considers all potential control conditions, thereby making best use of available data without the need to select just one particular condition as the control group. Second, it does not require any single control group to exhibit a parallel trend with the intervention group; rather, by construction, the control group matches the intervention group in *levels* of pre-policy outcomes.

The SC method requires a panel data structure with the same units of observation being followed over time. We therefore follow Abadie et al. (2010) and aggregate the patient level data to monthly proportions of same-day discharge at the level of the intervention group, i.e. one observation per month for each condition. We apply indirect standardisation to adjust for changes in case-mix over time. We estimate the relationship between patient characteristics and the probability of SDD in the financial year 2006 and then calculate an adjusted proportion of SDD for all months t

$$\hat{Y}_t = \frac{\sum_{i=1}^{N_t} Y_{it}}{\sum_{i=1}^{N_t} \hat{Y}_{it}} \times \bar{Y}_t \quad (8)$$

where \hat{Y}_{it} is the predicted probability of SDD for a patient in period t given the estimated relationship

$$Pr[Y_{i,2006}] = \alpha + \mathbf{X}'_{i,2006} \boldsymbol{\theta} + \epsilon_i \quad (9)$$

which we estimate as a linear probability model. This process is conducted separately for each condition. This approach assumes the relationship between patient characteristics and outcome is constant over time. Any deviations in adjusted predicted outcome between the intervention group and the control group can therefore be interpreted as improvements in the probability of SDD. Note that, as long as the same case-mix model is used for all periods, the choice of base year is arbitrary. Also, as our primary concern is changes over time, we do not include hospital fixed effects in predicting the proportion of SDD.

The pool of potential control units includes all non-incentivised conditions meeting the criteria of similar RR, admission pathway and minimum number of observations set out in Section 4.2. We specify the SC

algorithm to maximise similarity of the intervention and SC groups in terms of outcomes and average pre-policy patient characteristics. We then test if this is a reasonable control group from which to draw inferences by graphical assessment of how well the pre-policy outcomes of the intervention group are predicted by the control group, but also by constructing explicit tests. First, as recommended by Abadie et al. (2010), we assess goodness of fit in the pre-policy period by calculating the root mean square error (RMSE) of the predictions of the SDD rates of the control group compared to the intervention group for each pre-policy month. We reject the control group if the average RMSE exceeds 20% of the pre-policy SDD rate; similar to the $\pm 10\%$ rule used for selection DiD groups based on pre-trend. Second, good control groups should not consistently over- or under-predict the outcomes of the intervention group in the pre-policy period. We therefore construct a test statistic based on the number of times the monthly trend lines cross in the pre-policy period and reject control groups that cross less than 20% of the time. While these cut-offs are somewhat arbitrary, this procedure in effect operationalises the graphical analysis of the goodness of pre-treatment trajectories of the SC outcome.

The effect of the SDD BPT in year k is now given by

$$\tau_k = \frac{1}{12} \sum_{t=t^{BPT}+12(k-1)}^{t^{BPT}+12k} E[\hat{Y}_t^1 - \hat{Y}_t^0] \quad (10)$$

The average ATT over the post-policy period is computed as outlined in section 4.1.

As an SC model has a single treatment unit for each point in time, it is not appropriate to construct traditional standard errors. We therefore adopt the approach of placebo tests originally proposed by Abadie et al. (2010). We estimate a set of SC models, as described above, but treating each potential control unit as if it was the treated unit in turn. The original treatment unit is excluded from the set of potential controls for each of these placebo tests. From this process we acquire as many placebo ATT estimates as there are potential control units. We then apply the tests described above and drop any placebo results that do not meet the criteria.

The $\bar{\tau}^{Placebo}$ estimates from placebo tests represent variation in these values due to random chance. We therefore construct a p-value as the proportion of placebo tests with $|\bar{\tau}^{Placebo}| > |\bar{\tau}|$.¹⁴ We then convert these p-values to standard errors by using a normal approximation.

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

¹⁴ For example, if of 20 placebo tests, 3 have larger ATTs than the treatment model, this indicates $p=3/20=0.150$.

5. Results

We conduct ITS analysis for all 32 incentivised conditions. DiD and SC analyses are conducted for 18 of these conditions for which appropriate control groups are identified: 13 conditions are analysed using both DID and SC, 3 using just DID and 2 using just SC. We therefore discuss the ITS results and compare them with those supported by DID or SC analyses, where applicable. Time-series graphs of the proportion of patients being admitted and discharged on the same day with superimposed trend lines are presented in the online appendix.

5.1. Average effect over the post-policy period

Our main focus is on the average effect of the SDD BPT policy, represented by the parameter estimate $\bar{\tau}$ (i.e. the ATT in the average year over the full post-policy period) for each of the 32 conditions. Figures 2 and 3 present the estimated effects with 95% confidence intervals for incentivised planned and emergency conditions, respectively.

The results for the *planned* conditions generally support a (weakly) positive effect for 11 of the 13 conditions, with a weakly negative effect for two conditions (#9-10). However, the ITS effect is statistically significant for only four out of the 13 conditions (#1-4). The largest effects are for #1 cholecystectomy and #3 sentinel node mapping (>10pp), but are smaller (<5pp) for #2 simple mastectomy and #4 female incontinence management. The results from applying the DID and SC methods generally concur with ITS, with three exceptions that show positive and statistically significant effects (#5,7,13) under DID but not ITS (with >5pp for the latter two conditions).

The results for the *emergency* conditions in Figure 3 are more mixed, but tend to be of smaller magnitude than for the planned conditions. Of the 19 emergency conditions, ITS analysis indicates a non-significant effect for 12 conditions, while four have a significantly positive effect (#18-19,21,31) and three have a significantly negative effect (#14,20,24). The size of the effect ranges from -6pp to +6pp. For the 8 (out of 19) conditions for which a DID control group can be identified, the results from DID analysis generally concur with ITS, though now #15 acute headache appears significantly positive, #23 deep vein thrombosis appears significantly negative and #14 epilepsy is non-significant. For the 3 (out of 19) conditions for which a SC analysis is conducted, the effect is always very close to zero, including for two conditions estimated to have a significantly positive effect (#21, 31) and one estimated to have a significantly negative effect (#14) when applying alternative methods. The low number of SC analyses that satisfy our quality criteria is due to the limited pool of potential controls.

Taken together, these results indicate that the SDD pricing policy had a positive effect on planned conditions with a positive statistically significant effect for 4/13 conditions under ITS and 4/10 under DiD. The results are rather mixed for emergency conditions, with positive effects for 4/19 under ITS and 1/6 under DiD and negative effects for 3/19 under ITS and 2/6 under DiD. There is no general pattern to either the size of the mean effects or the relative widths of the confidence intervals when comparing the ITS and DiD results. The SC results appear to be more pessimistic compared to the other two methods.

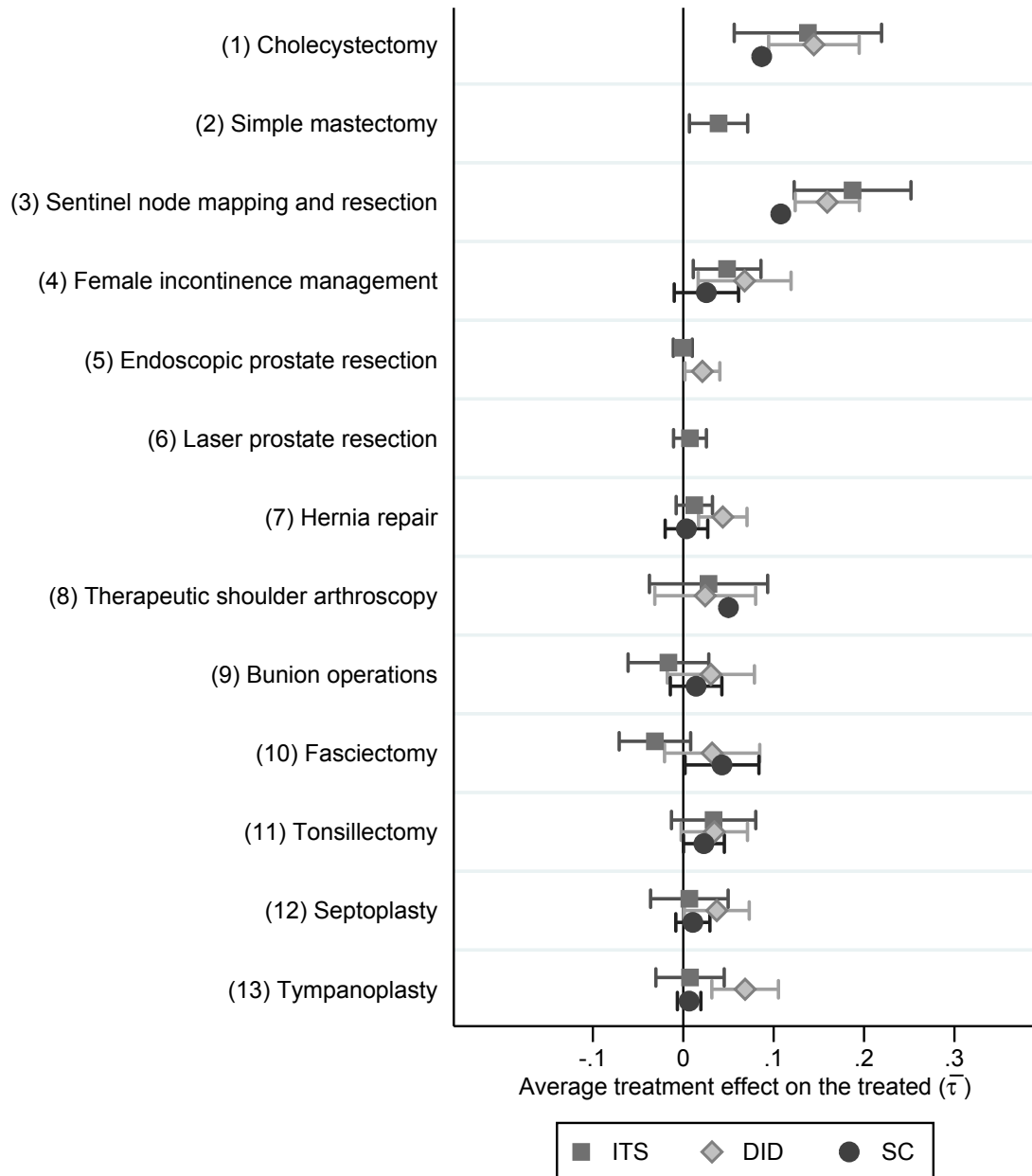


Figure 2: Average change in SDD rate over post-policy - planned conditions

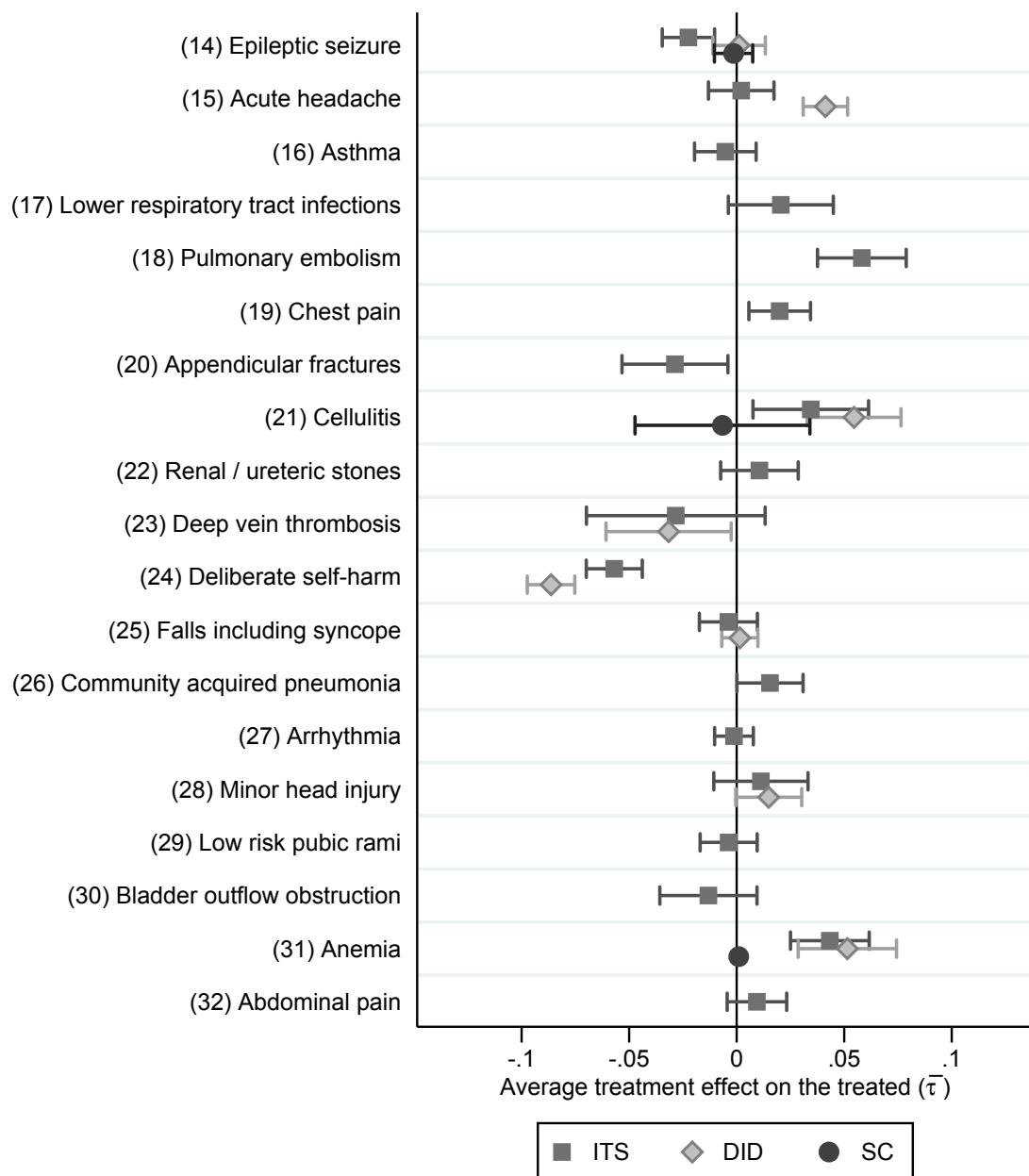


Figure 3: Average change in SDD rate over post-policy - emergency conditions

The significant ITS results translate into approximately 6,500 more patients admitted, treated and discharged on the same day in a year across all incentivised conditions¹⁵. As Figure 4 shows, these overall effects are driven by large positive effects for cholecystectomy (#1), sentinel node mapping (#3) and chest pain (#19), but these are offset by a large negative effect for self-harm (#24).

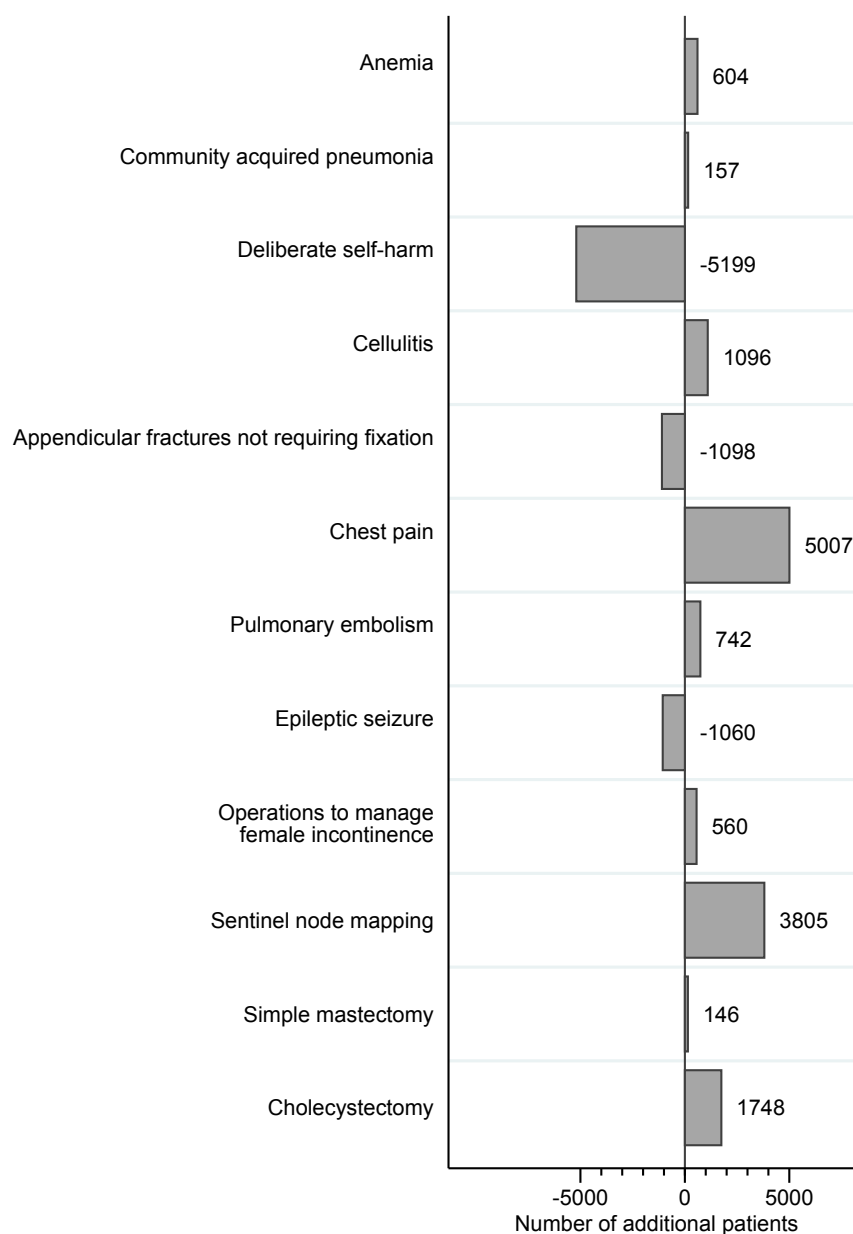


Figure 4: Additional SDD patients per year based on ITS estimates

¹⁵ The additional patients treated as SDD across all incentivised conditions in a given year is $\sum_{c=1}^{32} \bar{\tau} N_c$ where N_c is the number of patients within the scope of each incentivised condition c in the average post-policy year. Where the estimate for $\bar{\tau}$ is insignificant, we assume the value of this parameter is zero. Where the estimated effect is significant, we use the point estimate.

5.2. Time-varying effects

Our models allow for policy effects to vary over time and effects for each year after policy introduction are reported for each type of analysis in Tables 2-4. Focussing on the ITS results, as these are available for all 32 conditions, we find that 22 indicate at least one significant year effect. The patterns over time are non-linear and almost every possible combination of year-on-year effects is observed. We find conditions with initially positive and then strengthening effects (#1,3-4,18-19,21,31) or weakening effects (#7-8); and conditions with initially negative effects which grow more pronounced (#10,14,24) or less pronounced (#15-17,22-23,25,32). The results exhibit a similar variety of year-on-year patterns when we conduct DID (Table 3) or SC analysis (Table 4). The results suggest that there is no common behavioural response to the introduction of the SDD BPT over time.

5.3. Association with incentive design features

We now investigate if the response to policy is associated with features of the design of SDD incentives. 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} . To compare the estimated ITS effect across conditions we, therefore, compute the elasticities of the proportion of SDD with respect to price as

$$\epsilon = \frac{\bar{\tau} / \bar{Y}_{Pre}}{(P_1^{SDD} - P_1^{ON}) / P_1^{ON}} \quad (11)$$

where \bar{Y}_{Pre} is the observed outcome for the incentivised condition in the year before the announcement period. The median elasticity across the 13 planned and 19 emergency conditions is 0.24 and 0.01, respectively. Six conditions show an elasticity above 1.

Hospitals may respond more strongly for conditions offering relatively higher financial returns, keeping other factors constant. Figures 5a and 5b 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 5c 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. There is suggestive evidence that larger elasticities are concentrated in conditions with higher SDD prices and larger price differences. Moreover, elasticities appear to increase in the size of the total incentive $\Delta(P - AC) = (P_1^{SDD} - AC_0^{SDD}) - (P_1^{ON} - AC_0^{ON})$ and this association is more pronounced across planned SDD conditions.

We also explore whether responses appear to be driven by clinical reasons. We hypothesise that responses to the BPT 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 5d provides some support that larger elasticities occur for conditions with lower pre-policy SDD rates, but Figure 5e does not suggest a relationship between the elasticities and the gap between existing practice (i.e. pre-policy SDD rate) and the RR.

Table 2: Average treatment effect on the treated - ITS analyses

| # | Condition | Average ($\bar{\tau}$) | | Year 1 ($\tau_{k=1}$) | | Year 2 ($\tau_{k=2}$) | | Year 3 ($\tau_{k=3}$) | | Year 4 ($\tau_{k=4}$) | | Year 5 ($\tau_{k=5}$) | |
|----|--|--------------------------|----------|-------------------------|----------|-------------------------|----------|-------------------------|----------|-------------------------|----------|-------------------------|---------|
| | | Est | SE | Est | SE | Est | SE | Est | SE | Est | SE | Est | SE |
| 1 | Cholecystectomy | 0.138 | 0.042*** | 0.088 | 0.022*** | 0.121 | 0.033*** | 0.137 | 0.043** | 0.161 | 0.054** | 0.179 | 0.066** |
| 2 | Simple mastectomy | 0.039 | 0.016* | 0.017 | 0.009* | 0.048 | 0.019* | 0.048 | 0.020* | 0.043 | 0.022 | | |
| 3 | Sentinel node mapping | 0.187 | 0.033*** | 0.108 | 0.023*** | 0.164 | 0.030*** | 0.222 | 0.037*** | 0.234 | 0.043*** | | |
| 4 | Operations to manage female incontinence | 0.048 | 0.019* | 0.031 | 0.015* | 0.059 | 0.020** | 0.053 | 0.022* | 0.054 | 0.027* | | |
| 5 | Endoscopic prostate resection | -0.001 | 0.005 | 0.006 | 0.005 | 0.005 | 0.006 | 0.009 | 0.006 | -0.025 | 0.008** | | |
| 6 | Laser prostate resection | 0.007 | 0.009 | 0.000 | 0.007 | 0.006 | 0.009 | 0.009 | 0.011 | 0.016 | 0.013 | | |
| 7 | Hernia repair | 0.012 | 0.010 | 0.023 | 0.007** | 0.029 | 0.009** | 0.000 | 0.013 | -0.003 | 0.014 | | |
| 8 | Shoulder decompression | 0.028 | 0.033 | 0.050 | 0.020* | 0.062 | 0.029* | 0.010 | 0.040 | -0.006 | 0.048 | | |
| 9 | Bunion operation | -0.016 | 0.023 | 0.000 | 0.014 | 0.009 | 0.023 | -0.035 | 0.028 | -0.042 | 0.032 | | |
| 10 | Fasciectomy | -0.031 | 0.020 | 0.009 | 0.015 | -0.002 | 0.018 | -0.051 | 0.024* | -0.086 | 0.029** | | |
| 11 | Tonsillectomy | 0.034 | 0.024 | 0.023 | 0.019 | 0.032 | 0.025 | 0.045 | 0.030 | | | | |
| 12 | Septoplasty | 0.007 | 0.022 | 0.012 | 0.017 | 0.019 | 0.023 | -0.011 | 0.028 | | | | |
| 13 | Tympanoplasty | 0.007 | 0.019 | 0.009 | 0.017 | 0.006 | 0.023 | | | | | | |
| 14 | Epilepsy | -0.022 | 0.006*** | -0.016 | 0.005** | -0.029 | 0.007*** | -0.022 | 0.008** | | | | |
| 15 | Acute headache | 0.002 | 0.008 | -0.016 | 0.006* | 0.000 | 0.009 | 0.021 | 0.010* | | | | |
| 16 | Asthma | -0.005 | 0.007 | -0.017 | 0.007* | -0.003 | 0.008 | 0.004 | 0.009 | | | | |
| 17 | Respiratory | 0.021 | 0.012 | -0.004 | 0.012 | 0.023 | 0.014 | 0.044 | 0.016** | | | | |
| 18 | Pulmonary embolism | 0.058 | 0.011*** | 0.025 | 0.009** | 0.056 | 0.011*** | 0.093 | 0.014*** | | | | |
| 19 | Chest pain | 0.020 | 0.007** | 0.001 | 0.006 | 0.016 | 0.008 | 0.044 | 0.010*** | | | | |
| 20 | Appendicular fractures | -0.029 | 0.013* | -0.028 | 0.011* | -0.032 | 0.014* | -0.026 | 0.014 | | | | |
| 21 | Cellulitis | 0.034 | 0.014* | 0.004 | 0.011 | 0.031 | 0.015* | 0.066 | 0.020*** | | | | |
| 22 | Renal/ureteric stones | 0.011 | 0.009 | -0.002 | 0.008 | 0.006 | 0.010 | 0.028 | 0.013* | | | | |
| 23 | Deep vein thrombosis | -0.028 | 0.021 | -0.064 | 0.017*** | -0.034 | 0.023 | 0.011 | 0.027 | | | | |
| 24 | Deliberate self-harm | -0.057 | 0.007*** | -0.047 | 0.005*** | -0.061 | 0.007*** | -0.063 | 0.009*** | | | | |
| 25 | Falls | -0.004 | 0.007 | -0.017 | 0.006** | -0.006 | 0.008 | 0.013 | 0.010 | | | | |
| 26 | Pneumonia | 0.015 | 0.008* | 0.006 | 0.008 | 0.024 | 0.009* | | | | | | |
| 27 | Fibrillation | -0.001 | 0.005 | -0.008 | 0.004 | 0.005 | 0.006 | | | | | | |
| 28 | Head injury | 0.011 | 0.011 | 0.006 | 0.011 | 0.017 | 0.013 | | | | | | |
| 29 | Pelvis fracture | -0.004 | 0.007 | -0.004 | 0.007 | -0.004 | 0.008 | | | | | | |
| 30 | Bladder outflow | -0.013 | 0.012 | -0.016 | 0.011 | -0.010 | 0.015 | | | | | | |
| 31 | Anaemia | 0.043 | 0.009*** | 0.022 | 0.009* | 0.064 | 0.011*** | | | | | | |
| 32 | Abdominal pain | 0.009 | 0.007 | -0.004 | 0.006 | 0.023 | 0.008** | | | | | | |

*** p< 0.001; ** p<0.01; * p<0.05

Standard errors (SEs) are clustered at hospital level.

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

| # | Condition | Average ($\bar{\tau}$) | | Year 1 ($\tau_{k=1}$) | | Year 2 ($\tau_{k=2}$) | | Year 3 ($\tau_{k=3}$) | | Year 4 ($\tau_{k=4}$) | | Year 5 ($\tau_{k=5}$) | |
|----|--|--------------------------|----------|-------------------------|----------|-------------------------|----------|-------------------------|----------|-------------------------|----------|-------------------------|-------|
| | | Est | SE | Est | SE | Est | SE | Est | SE | Est | SE | Est | SE |
| 1 | Cholecystectomy | 0.145 | 0.026*** | 0.085 | 0.024*** | 0.120 | 0.028*** | 0.133 | 0.031*** | 0.177 | 0.028*** | 0.205 | 0.001 |
| 3 | Sentinel node mapping | 0.159 | 0.018*** | 0.084 | 0.019*** | 0.139 | 0.019*** | 0.196 | 0.020*** | 0.198 | 0.023*** | | |
| 4 | Operations to manage female incontinence | 0.068 | 0.026** | 0.038 | 0.024 | 0.072 | 0.027** | 0.094 | 0.030** | 0.071 | 0.031* | | |
| 5 | Endoscopic prostate resection | 0.021 | 0.010* | 0.021 | 0.010* | 0.014 | 0.012 | 0.023 | 0.011* | 0.027 | 0.012* | | |
| 7 | Hernia repair | 0.044 | 0.014** | 0.031 | 0.015* | 0.055 | 0.015*** | 0.033 | 0.019 | 0.056 | 0.019** | | |
| 8 | Shoulder decompression | 0.024 | 0.028 | -0.001 | 0.027 | 0.017 | 0.033 | 0.028 | 0.036 | 0.049 | 0.032 | | |
| 9 | Bunion operation | 0.031 | 0.025 | 0.001 | 0.024 | 0.018 | 0.029 | 0.038 | 0.031 | 0.070 | 0.029* | | |
| 10 | Fasciectomy | 0.032 | 0.027 | 0.025 | 0.026 | 0.022 | 0.031 | 0.038 | 0.033 | 0.044 | 0.031 | | |
| 11 | Tonsillectomy | 0.034 | 0.019 | 0.031 | 0.020 | 0.023 | 0.022 | 0.049 | 0.021* | | | | |
| 12 | Septoplasty | 0.037 | 0.018* | 0.006 | 0.018 | 0.048 | 0.021* | 0.056 | 0.021** | | | | |
| 13 | Tympanoplasty | 0.068 | 0.019*** | 0.054 | 0.020** | 0.084 | 0.022*** | | | | | | |
| 14 | Epilepsy | 0.001 | 0.006 | 0.009 | 0.009 | 0.002 | 0.008 | -0.008 | 0.009 | | | | |
| 15 | Acute headache | 0.041 | 0.005*** | -0.017 | 0.005** | 0.148 | 0.007*** | -0.010 | 0.006 | | | | |
| 21 | Cellulitis | 0.055 | 0.011*** | 0.027 | 0.012* | 0.060 | 0.014*** | 0.075 | 0.015*** | | | | |
| 23 | Deep vein thrombosis | -0.032 | 0.015* | -0.056 | 0.015*** | -0.041 | 0.017* | 0.000 | 0.018 | | | | |
| 24 | Deliberate self-harm | -0.086 | 0.006*** | -0.072 | 0.006*** | -0.088 | 0.006*** | -0.100 | 0.007*** | | | | |
| 25 | Falls | 0.001 | 0.004 | -0.009 | 0.005 | 0.006 | 0.005 | 0.008 | 0.005 | | | | |
| 28 | Head injury | 0.015 | 0.008 | 0.017 | 0.009* | 0.012 | 0.009 | | | | | | |
| 31 | Anaemia | 0.051 | 0.012*** | 0.040 | 0.012*** | 0.062 | 0.014*** | | | | | | |

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Standard errors (SEs) are clustered at hospital level.

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

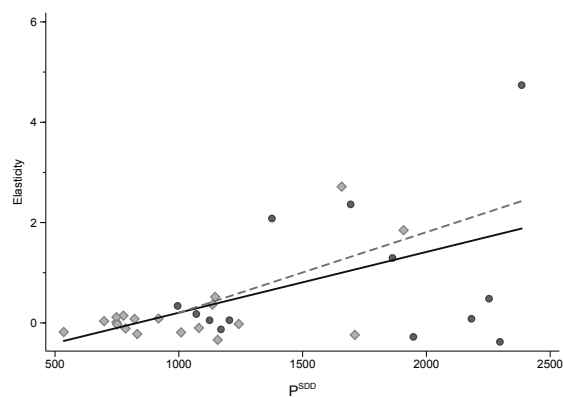
| # | Condition | Average ($\bar{\tau}$) | | Year 1 ($\tau_{k=1}$) | | Year 2 ($\tau_{k=2}$) | | Year 3 ($\tau_{k=3}$) | | Year 4 ($\tau_{k=4}$) | | Year 5 ($\tau_{k=5}$) | | Number of placebo tests |
|----|--|--------------------------|--------|-------------------------|--------|-------------------------|--------|-------------------------|--------|-------------------------|-------|-------------------------|-------|-------------------------|
| | | Est | SE | Est | SE | Est | SE | Est | SE | Est | SE | Est | SE | |
| 1 | Cholecystectomy | 0.087 | †*** | 0.077 | †*** | 0.107 | †*** | 0.135 | †*** | 0.173 | †*** | 0.190 | 0.104 | 15 |
| 3 | Sentinel node mapping | 0.108 | †*** | 0.091 | †*** | 0.144 | †*** | 0.203 | †*** | 0.194 | †*** | | | 63 |
| 4 | Operations to manage female incontinence | 0.026 | 0.018 | 0.020 | 0.021 | 0.065 | 0.043 | 0.070 | 0.046 | 0.084 | 0.045 | | | 63 |
| 7 | Hernia repair | 0.004 | 0.012 | 0.003 | 0.026 | 0.008 | 0.025 | 0.007 | 0.032 | 0.012 | 0.045 | | | 82 |
| 8 | Shoulder decompression | 0.050 | †*** | 0.058 | 0.033 | 0.085 | 0.040* | 0.058 | 0.046 | 0.066 | 0.043 | | | 63 |
| 9 | Bunion operation | 0.014 | 0.015 | 0.027 | 0.024 | 0.052 | 0.038 | 0.016 | 0.039 | 0.020 | 0.045 | | | 82 |
| 10 | Fasciectomy | 0.043 | 0.021* | 0.078 | 0.031* | 0.114 | †*** | 0.107 | 0.052* | 0.110 | 0.057 | | | 75 |
| 11 | Tonsillectomy | 0.023 | 0.012* | 0.054 | 0.029 | 0.059 | 0.032 | 0.078 | 0.039* | | | | | 63 |
| 12 | Septoplasty | 0.011 | 0.010 | 0.028 | 0.026 | 0.036 | 0.031 | 0.027 | 0.037 | | | | | 63 |
| 13 | Tympanoplasty | 0.006 | 0.007 | 0.028 | 0.028 | 0.040 | 0.037 | | | | | | | 64 |
| 14 | Epilepsy | -0.001 | -0.005 | 0.005 | 0.015 | -0.003 | ‡ | -0.014 | -0.043 | | | | | 4 |
| 21 | Cellulitis | -0.007 | -0.021 | -0.029 | -0.025 | -0.032 | -0.047 | 0.003 | 0.010 | | | | | 4 |
| 31 | Anaemia | 0.001 | ‡ | -0.008 | ‡ | 0.015 | 0.046 | | | | | | | 4 |

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

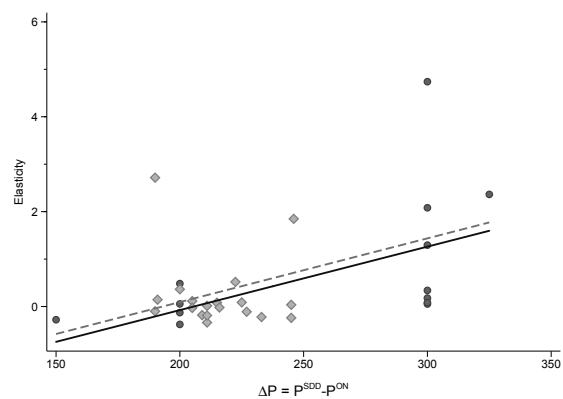
Standard errors (SEs) are obtained through p-value inversion, where p-values are calculated using the placebo test proposed by Abadie et al. (2010).

† All placebo tests generate $|\bar{\tau}^{Placebo}| < |t_{\tau}|$ so that no SE can be calculated.

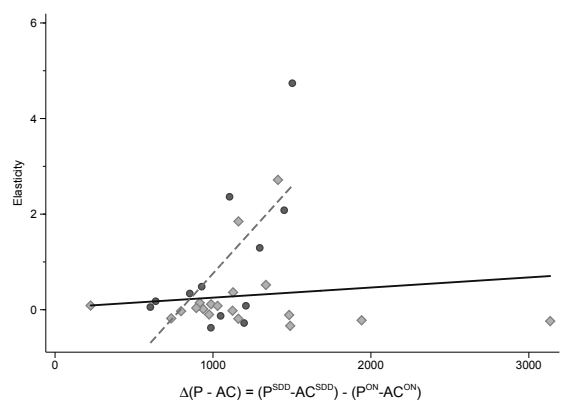
‡ No placebo group fulfilled our minimum quality criteria for a good SC group (see Section 4.3).



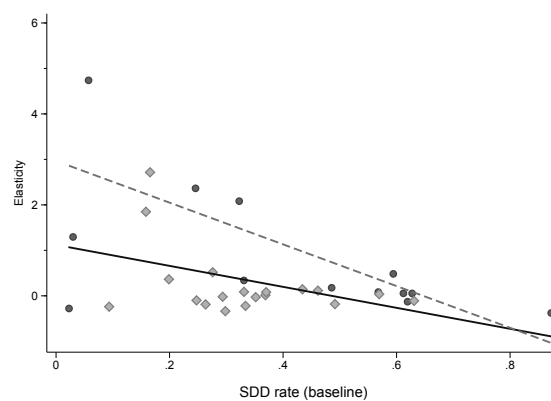
(a) SDD price



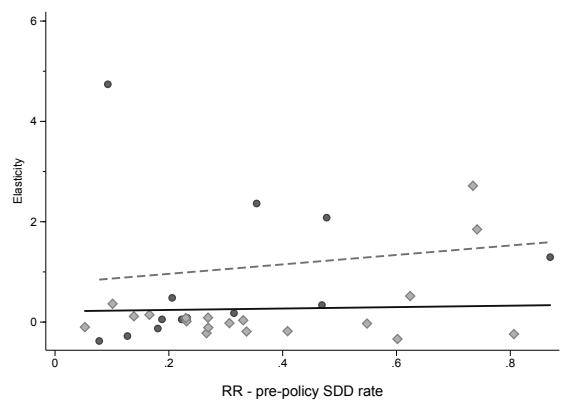
(b) % price difference SDD vs. ON



(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 5: Association between price elasticity of SDD care and tariff design factors

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 costs and are also higher than the price for otherwise identical hospital care that involves an overnight stay. This P4E policy is, therefore, unusual both in having different objectives to most P4P schemes and also in offering high-powered incentives.

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). This supported the roll-out of the policy to 31 more conditions. Our study set out to assess how far the original findings are generalisable and would also be observed for 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. Evaluating across all 32 conditions, we do find a positive response, translating into approximately 6,500 more patients treated on an SDD basis per year. However, perhaps surprisingly, we do not find a consistent positive response across all incentivised conditions. Indeed, for some 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 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 such as those admitted late in the evening. This may limit the scope for rapid increases in SDD rates in emergency conditions compared to planned conditions that permit efficient scheduling.

It has been argued that the limited impact of P4P schemes is due to incentives being too small and the incentivised behaviour lacking clinical buy-in. 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; in effect evaluating 32 separate experiments. 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 but there is a positive association, especially for planned conditions, when both price and cost advantages of SDD care are taken into consideration.

In conclusion, we find some evidence that hospitals respond to price signals and that payers, therefore, can use pricing instruments to improve supply-side 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 control 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.

References

- Abadie, A. and G. Imbens (2006). "Large Sample Properties of Matching Estimators for Average Treatment Effects". *Econometrica*, 74: 235–267.
- 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.
- 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.
- Allin, S., M. Baker, M. Isabelle, and M. Stabile (2015). *Physician Incentives and the Rise in C-sections: Evidence from Canada*. Working Paper 21022. National Bureau of Economic Research. DOI: 10.3386/w21022. URL: <http://www.nber.org/papers/w21022>.
- BAAEC (2014). *Directory of Ambulatory Emergency Care for Adults*. Vol. 4.
- BADS (2006). *BADS Directory of Procedures*. 3rd ed. London.
- Bharadwaj, P., J. Johnson, and K. Løken (2014). "Smoking bans, maternal smoking and birth outcomes". *Journal of Public Economics*, 115: 72–93.
- 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.
- 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.
- 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*.
- (2010). *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.
- Foo, P., R. Lee, and K. Fong (2017). "Physician Prices, Hospital Prices, and Treatment Choice in Labor and Delivery". *American Journal of Health Economics*, 3: 422–453.
- 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: 50.

- Green, C., J. Heywood, and M. Navarro (2014). "Did liberalising bar hours decrease traffic accidents". *Journal of Health Economics*, 35: 189–198.
- Gruber, J., J. Kim, and D. Mayzlin (1999). "Physician Fees and Procedure Intensity: The Case of Cesarean Delivery". *Journal of Health Economics*, 18: 473–490.
- Hafsteinsdottir, E. and L. Siciliani (2010). "DRG prospective payment systems: refine or not refine?" *Health Economics*, 19: 1226–1239.
- Hodgkin, D. and T. McGuire (1994). "Payment levels and hospital response to prospective payment". *Journal of Health Economics*, 13: 1–29.
- 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.
- 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.
- 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 and NHS England (2014). *2014/15 National Tariff Payment System*.
- Papanicolas, I. and A. McGuire (2015). "Do financial incentives trump clinical guidance? Hip Replacement in England and Scotland." *Journal of Health Economics*, 44: 25–36.
- 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.
- 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*.
- Rubin, D. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies". *Journal of Educational Psychology*, 66: 688–701.
- Shleifer, A. (1985). "A Theory of Yardstick Competition". *RAND Journal of Economics*, 16: 319–327.
- 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.

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 | 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 | Shoulder decompression | 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 | Epilepsy | 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 | Respiratory | 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 | 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 | 67.6 | 0.52 | 0.16 | 2.46 | 0.37 |
| 26 | Pneumonia | 51.8 | 0.50 | 0.16 | 0.63 | 0.22 |
| 27 | Fibrillation | 68.1 | 0.48 | 0.14 | 3.42 | 0.39 |
| 28 | Head injury | 54.9 | 0.56 | 0.18 | 1.63 | 0.33 |
| 29 | Pelvis fracture | 81.3 | 0.15 | 0.14 | 2.43 | 0.37 |
| 30 | Bladder outflow | 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 | 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 | Shoulder decompression | 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 | Epilepsy | 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 | Respiratory | 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 | 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 | 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 | 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 | Fibrillation | 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 | 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 | Pelvis fracture | 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 | 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% |