CMPO Working Paper Series No. 08/205

Incentives and Targets in Hospital Care: Evidence from a Natural Experiment

Carol Propper[1,2,3] **Matt Sutton**[4,5]

Carolyn Whitnall[1] **Frank Windmeijer**[1,2]

[1] Centre for Market and Public Organisation, University of Bristol
[2] Department of Economics, University of Bristol
[3] Imperial College, London and Centre for Economic Policy Research, London
[4] Health Economics, Community-Based Medicine, University of Manchester
[5] Health Economics Research Unit, University of Aberdeen

October 2008

Published in Journal of Public Economics 94 (3-4) pp. 318-335 (April 2010)

Abstract

Performance targets are commonly used in the public sector, despite their well known problems when organisations have multiple objectives and performance is difficult to measure. It is possible that such targets may work where there is considerable consensus that performance needs to be improved. We investigate this possibility by examining the response of the English National Health Service (NHS) to waiting time targets. Long waiting times have been a key issue for the NHS for many years. Using a natural policy experiment exploiting differences between countries of the UK, supplemented with a panel of data on English hospitals, we examine whether high profile targets to reduce waiting times met their goals of reducing waiting times without diverting activity from other less well monitored aspects of health care. Using this robust design, we find that targets led to a fall in waiting times without apparent reductions in other aspects of patient care.

Keywords: health care, waiting times, targets, incentives

JEL Classification: I18, L32

Electronic version: www.bristol.ac.uk/cmpo/publications/papers/2008/wp205.pdf

Acknowledgements

We thank those at ISD Scotland and the Department of Health who provided us with data, in particular David Bailey for preparing the samples from HES and SMR01 and Stephen Young for assembling the Scottish waiting list census data. This project received funding from the ESRC under the Public Services Delivery Programme, grant no. RES-166-25-0022, and the ESRC Centre Grant to CMPO, grant no. RES-343-28-3001.

Address for Correspondence
CMPO, Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol
BS8 1TX
Carol.Propper@bristol.ac.uk
www.bristol.ac.uk/cmpo/

CMPO is jointly funded by the Leverhulme Trust and the ESRC

CMPO Working Paper Series No. 08/205

CMPO is jointly funded by the Leverhulme Trust and the ESRC

CMPO Working Paper Series No. 08/205

Incentives and Targets in Hospital Care: Evidence from a Natural Experiment

Carol Propper[1,2,3] **Matt Sutton**[4,5]

Carolyn Whitnall[1] **Frank Windmeijer**[1,2]

[1] *Centre for Market and Public Organisation, University of Bristol*
[2] *Department of Economics, University of Bristol*
[3] *Imperial College, London and Centre for Economic Policy Research, London*
[4] *Health Economics, Community-Based Medicine, University of Manchester*
[5] *Health Economics Research Unit, University of Aberdeen*

October 2008

Abstract

Performance targets are commonly used in the public sector, despite their well known problems when organisations have multiple objectives and performance is difficult to measure. It is possible that such targets may work where there is considerable consensus that performance needs to be improved. We investigate this possibility by examining the response of the English National Health Service (NHS) to waiting time targets. Long waiting times have been a key issue for the NHS for many years. Using a natural policy experiment exploiting differences between countries of the UK, supplemented with a panel of data on English hospitals, we examine whether high profile targets to reduce waiting times met their goals of reducing waiting times without diverting activity from other less well monitored aspects of health care. Using this robust design, we find that targets led to a fall in waiting times without apparent reductions in other aspects of patient care.

Keywords: health care, waiting times, targets, incentives

JEL Classification: I18, L32

Electronic version: www.bristol.ac.uk/cmpo/publications/papers/2008/wp205.pdf

Acknowledgements

We thank those at ISD Scotland and the Department of Health who provided us with data, in particular David Bailey for preparing the samples from HES and SMR01 and Stephen Young for assembling the Scottish waiting list census data. This project received funding from the ESRC under the Public Services Delivery Programme, grant no. RES-166-25-0022, and the ESRC Centre Grant to CMPO, grant no. RES-343-28-3001.

Address for Correspondence
CMPO, Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol
BS8 1TX
Carol.Propper@bristol.ac.uk
www.bristol.ac.uk/cmpo/

CMPO is jointly funded by the Leverhulme Trust and the ESRC

Introduction

The private sector has long linked performance to rewards. Governments seeking to drive up productivity in the public sector have sought to emulate this (Osborne and Gaebler, 1992). Performance has been linked directly to financial rewards: examples include performance based pay for managers and teams in the Job Training and Partnership Act of 1983 (Heckman et al., 1997) and performance related pay for teachers in Israel (Lavy, 2008). More commonly, performance of public organisations has been made explicit and organisations have been set targets for performance. In this case, the link to rewards is less direct than in performance related pay but operates through career concerns of senior managers.

This interest in linking performance and reward is despite theoretical analyses which suggest that the public sector may be particularly susceptible to many of the pitfalls of formal incentives. The multiple objectives that many public agencies face give agents opportunities to divert activity away from non-incentivised tasks, so that the optimal contract may have very low powered incentives (Holmstrom and Milgrom, 1991; Baker, 1992). The inherent inefficiency of the public sector may mean that monitoring for proper incentives may be weak (Prendergast, 2003), so giving agents opportunities to game the performance targets set by their political masters (for example, Courty and Marschke, 1997; Bevan and Hood 2006a; Smith, 1995).

However, from another perspective, the use of high profile targets accompanied by monitoring of agents and publication of performance may be beneficial. High profile targets may act as missions around which employees may coalesce (Friedman and Kelman, 2007). Dewatripont et al. (1999) have stressed the importance of missions in increasing productivity in public sector organisations. Besley and Ghatak (2003, 2005) have shown that missions, through the associated matching of mission orientated firms and workers, can be a substitute for explicit financial rewards. The link to rewards may also decrease the amount of effort spent on tasks which public employees value but bring less social welfare (e.g. Heckman et al., 1997). So a target directed towards performance improvement, where performance is widely acknowledged to be in need of enhancement, may improve the measured outcome without diversion of activity or gaming.

This paper investigates this idea by examining the responses to a high profile target regime implemented in a large public sector organisation. In 2000, in response to widespread dissatisfaction with waiting times for hospital care, the English government instituted an aggressive target based policy to reduce the very long waiting lists for non-emergency care in the National Health Service (NHS).[1] Target maximum waiting times were set on an annual basis, monitored monthly and reduced each year. NHS hospitals in England were to have no patients waiting for inpatient treatment for more than 18 months by the end of March 2001, a length of time that was to decrease annually by 3 months until a maximum of six months in December 2005. Performance against targets was published widely and used as the basis for direct sanctions and rewards. Managers of poorly performing hospitals could be fired, while managers of high performing hospitals were granted greater autonomy in how they managed their hospitals and the freedom to keep certain surpluses. This policy was unprecedented in three ways; it was long-term, specified escalating targets, and was rigorously enforced and monitored. The penalties for managerial failure in meeting these targets were judged so strong that the regime has been dubbed one of ‘targets and terror’ and likened to the targets set for managers of state enterprises in pre-reform Soviet Russia (Hood and Bevan, 2005; Bevan and Hood, 2006b).

Hospitals are multi-product organisations, so targets on waiting lists could lead to reductions of effort on less well monitored outputs. In addition, the performance measure – an absolute number of people waiting less than a certain amount of time - may have been manipulated by managers to give the appearance of hitting the target, whilst in reality not achieving them. This raises the concern - widely expressed by the public, political commentators and clinicians when the targets were announced (Kelman and Friedman, 2007) - that the use of a high powered incentive system to reduce waiting times might have ‘hit the target but missed the point’.

This paper evaluates whether this was the case. We examine first whether the target regime achieved its goal of reducing the long waits for elective care. Second, we examine whether this was at the expense of performance on other activities which were not subject to targets. Third, we ask whether hospitals ‘gamed’ the targets, by categorising patients in ways that meant they were not counted or by reshuffling patients on the list so that patients were treated

[1] The reduction of waiting times targets was one of the five election pledges used by the Blair campaign to win the 1997 general election.

in terms of list priority rather than medical need. Finally, we examine the impact of the policy on quality of patient care.

To do this we exploit the fact that the targets were a natural experiment. The policy was implemented in England. Pre-1999 health policy was common to all four countries of the United Kingdom (UK) and set in Westminster. Post-1999 the UK government devolved responsibility over various domains of public policy to each country, including responsibility for the health service. None of the other countries in the UK chose to implement the waiting list target policy. We use this difference between England and Scotland, the largest of the three other countries in the UK, to identify the impact of the target regime[2]. We supplement this difference-in-difference approach with analyses at the hospital level for England to test that the responses to pressure from waiting times targets that we observe at country level are mirrored at the hospital level.

We find that the targets reduced waiting times by 13 days at the mean, with considerably larger reductions at the top end of the distribution. Levels of non-emergency (elective) care rose, with no apparent reductions in non-targeted activity (emergency care and length of stay). We find no evidence of re-ordering of patients on lists to meet targets. Nor do we find evidence of a fall in patient quality: in fact, in terms of some outcomes, we find that quality of care in England rose post policy. We do, however, find some evidence of waiting list manipulation: patients were removed, temporarily and permanently, from waiting lists. On the basis of the quality measures we examine, we conclude that these suspensions and removals were not harmful to patient health.

1. Institutional background and possible responses to targets

1.1 The institutional background

Health care in the UK is predominantly provided by the National Health Service (NHS), which is funded by general taxation, free to the consumer at the point of use and employs

[2] We select Scotland, rather than Northern Ireland or Wales, as the appropriate comparator to England. It is the largest of the three devolved administrations and has a greater degree of devolution and independence of its Parliament. It also has a more self-contained healthcare system with less cross-border flows between England and Scotland as compared to between England and Wales.

around 1.2 million people.[3] In the NHS, purchasing organisations receive budgets from central government to procure care for geographically-defined populations. In the period we examine, these purchasers negotiated contracts with local groupings of hospitals, known as NHS Trusts (referred to as hospitals in the rest of this paper). Three-quarters of contracts between purchasers and these providers were ‘block’ contracts, in which payment was only linked to volume in cases of extreme under- or over-performance (Goddard et al., 1997). For most purchasers there was a high degree of concentration of business with their local suppliers. On average, the main local provider accounted for nearly 70 percent of a purchaser’s admissions in the financial year 2002 (Dusheiko et al., 2008). Competition between hospitals was not encouraged in the period studied here.

For emergency care, patients have direct access to specialist treatment. For all other hospital care (known as elective care), which accounts for around half of all care, they must first contact their family doctor (GP). The GP provides a referral to a specialist employed in an NHS hospital. The individual waits for this first specialist appointment and, if more intensive treatment is required, then waits again for admission to hospital. The long periods for which patients had to wait for treatment were a very sensitive political issue, particularly the waiting times between seeing a specialist and admission to hospital, known as the inpatient waiting time. The targets to reduce these waiting times is our focus here.

Prior to devolution in 1999 inpatient waiting time targets were set by the Patients’ Charter (Department of Health, 1995). The Charter was common to England and Scotland and set a maximum waiting time of 18 months for 1997 and 1998 for England and 12 months for Scotland. It was not backed up by strong managerial sanctions. In 1999, the UK government devolved responsibility for the health service, creating new administrations with responsibility for policy and provision of NHS care in Scotland, Wales and (for some time) Northern Ireland.

Post this devolution the Department of Health in England in 2000 announced an ambitious ‘modernisation and reform’ programme (Department of Health, 2000). The ‘target and terror’ regime for the time spent waiting for inpatient treatment that is outlined above was a key

[3] There is a limited private sector which specialises in treatments for which there are long waiting lists. Demand in this sector has been shown to be a function of NHS waiting lists (Besley et al., 1999). Over the period analysed here demand for private care remained relatively static (Laing, 2007).

plank of these reforms. The Scottish Executive chose not to adopt this regime. Instead, from devolution in 1999, it focused on the abolition of the 1990s 'quasi-market' and the re-introduction of a professionally-led, integrated system based on concepts such as managed clinical networks (Alvarez-Rosete et al., 2005). Targets played little role. The overall guarantees of 12 months remained, though there were (new) exceptions for certain (undefined) conditions of 'low clinical priority.' In 2000 the Scottish Executive set down an "expectation" that waiting times should not exceed 9 months by the end of December 2003 (Scottish Executive, 2000) but waiting times at hospital level were not made public and this expectation was not strongly monitored. There were some signs that the policy in Scotland changed in 2003 when the White Paper issued in February 2003 (Scottish Executive, 2003) offered patients "a guarantee that our national targets will be met…[and] monitored". In November 2003 the Scottish Executive refined its objectives to "12 National Priorities", one of which was reducing waiting times. However, this greater focus was still not accompanied by publication of performance at hospital level or the coupling of performance against targets and managerial sanctions that operated in England.

Table 1 summarises the differences in the maximum waiting times set for hospitals in each year.[4] The level in England was higher initially. The announcement in 2000 - which covered the time period up to 2005 - was to drive the maximum waiting time steadily downwards in England through a set of ever-stricter targets. The Scottish administration maintained their target of 12 months until 2003/4 and only then lowered it. The net effect was that by 2004/5 the target in England was below that allowed in Scotland despite being 6 months higher in 2000.

Table 1. Target maximum waiting times (months) in England and Scotland

	97/98	**98/09**	**99/00**	**00/01**	**01/02**	**02/03**	**03/04**	**04/05**
England	18	18	18	18	15	12	9	6
Scotland	12	12	12	12	12	12	9	9

Source: Propper et al. (2008b). All targets were enforced from the end of the fiscal year (March), with the exception of the 6 months target, which applied from Dec 2005.

[4] Purchasers were also assessed on whether the waiting times of their residents at local hospitals complied with the target but had little influence over this performance.

1.2 The impact of targets on managerial behaviour in English hospitals

Targets were set on the stock of patients on the elective admission list. A hospital manager facing the target could take three types of action. She could increase effort to reduce the number who would wait longer than the prescribed length of time (the monitored task) and could divert effort away from non-monitored tasks. But on top of this, as the performance measure is a target, the manager could 'game' the target by taking actions which make it seem that the target has been met when in fact it has not.[5]

Effort on the monitored task: We assess the extent of effort on the monitored task by examining the waiting times of treated patients. This was the object of the policy. However, in order to check that managers did this without reducing the number of people treated, we also examine the number of elective patients treated and the number of patients put on to the waiting list[6].

Diversion of effort from non-monitored tasks: We assess the extent of possible effort diversion from non-monitored tasks by looking at the volume of emergency patients and the length of stay of elective patients. Effort could be diverted away from the treatment of emergency patients, though as hospitals have little control over the inflow of patients attending emergency departments we might expect relatively little reduction in the volumes of such patients. Less effort spent on elective patients who were admitted is more likely. For example, hospitals could reduce bed utilisation by increasing the proportion of patients treated with no overnight stay (day cases) and by decreasing lengths of stay. This would leave more beds available to treat those patients still on the waiting list, thereby achieving lower waiting times.

We examine two aspects of 'gaming': re-classification and re-prioritisation. Re-classification occurred if patients waiting for elective care were reclassified in a way that meant their waiting time was not counted. Re-prioritisation occurred if patients on the waiting lists were

[5] Gaming is a term widely used in the literature on responses to targets (Propper and Wilson, 2003). Here we use it to refer to actions which are designed to give the appearance of meeting the target whilst in practice not doing so or doing so at a cost to some of those on the list.

[6] There is little that hospitals can do to reduce the demand for referrals from GPs to see their specialists but they have control over the proportion of these added to the waiting list for hospital admission. Since the same specialists assess patients referred by GPs in outpatient clinics and treat patients during admission events, there may be strong substitution between them.

reshuffled so that those most likely to breach the target were treated first. Both were alleged to have occurred widely.[7]

Reclassification: There were in fact two lists for elective patients post-referral by a specialist. The first was the 'active list' of patients who were deemed to be ready for surgery. The second was the inpatient 'deferred list' for patients who needed treatment but who, for personal or medical reasons, were not yet in a position to have it. This designation was decided by hospital managers and clinicians. Only the 'active list' was subject to targets and this gave opportunities for reclassification. Three categories of elective patients were not part of the 'active list': planned admissions, suspensions and removals. Planned admissions are those which are scheduled according to clinical factors (e.g. a course of chemotherapy or the second of two hip replacements). These admissions were not covered by the waiting time targets. Suspended patients are those patients who were deemed either not to be medically ready for treatment or could not attend when first given an appointment date. They were put onto the 'deferred list' and not counted towards the target. Subsequently, when they were returned to the 'active list' their waiting times were calculated excluding the time spent suspended. Patients could also be removed from the list (for example, if they were on the list and then subsequently died, or were treated in the community or in another hospital). Removals were not counted in waiting list statistics.

Each of these three patient categories offered managers ways to manipulate performance against the targets. The categorisation of patients as planned is less likely to be used by managers in response to current pressure, as patients were classified as planned at the point when they were added to the list. Such re-categorisation would only reduce future pressure and would have little impact on current pressure from targets. On the other hand, patients could be suspended whilst they were on the 'active' list by transferring them to the 'deferred list' and patients who were removed were simply taken off the 'active' list. Therefore both suspensions and removals could be used to reduce the probability that the hospital would breach its target in the current quarter.

[7] For example, in the court case of Henry versus the British Broadcasting Corporation it was alleged that the following actions took place at the instigation of senior managers: removal of patients at or near the 18-month target from the inpatient list onto the deferred list: deliberately not admitting patients so they would not show on the inpatient list at the end of the month (the census date) and giving priority to long wait 'routine' patients over patient classified as 'soon' to ensure that long waiters would be admitted sooner (EWHC 2787 (QB), 2006).

Re-prioritisation: The targets gave strong incentives to reshuffle patients on the waiting list to treat those who were most likely to breach the target.[8] While NHS policy stressed that the order in which patients on lists should be treated should be determined by clinical priority only, there was a widespread belief amongst clinicians that waiting lists altered priorities away from those most in need to those with less urgent need for care (National Audit Office, 2001). We examine whether waiting lists were re-prioritised by analysing the proportion of patients admitted within 14 days for types of treatments that were admitted urgently prior to the introduction of the targets. We also examine whether more patient were admitted with complications under the target regime. If patients who needed medical care were being denied treatment in favour of less medically needy cases, the knock-on effect might be that when those patients were finally admitted they were in a worse state of health.

Outcomes: Ultimately what matters is whether any apparent effort diversion or gaming results in lower quality of care. We therefore assess whether these actions affected the quality of patient care. We do this by examining the impact of the policy on a set of measures of patient mortality. These outcome measures were chosen because mortality rates are often used to assess hospital quality, both by the regulatory agencies including UK government and in the research literature (e.g. Gaynor, 2006; Kessler and McClellan, 2000; Propper et al., 2008a).

2. Methodology and Data

2.1 Difference-in-difference methodology

Our primary identification strategy is to exploit the natural experiment. We use a difference-in-difference methodology at the country level and estimate:

(1) $$o_{jt} = \alpha + \phi t + \beta E_j + \gamma I_{[t\in pr]} + \delta I_{[t\in pr]} E_j + x'_{jt}\theta + \varepsilon_{jt}$$

where o_{jt} is the outcome of interest for country $j =$ Scotland or England at time t; $E_j = 1$ denotes England; $I_{[t\in pr]} = 1$ if the period t is during the policy regime (2000/1 onwards) and

[8] Where targets were fixed over time one response is to 'stack up' patients just below the target waiting time. In the environment we study, where targets were monitored monthly and reduced every year, stacking up patients just below the current year's target is less helpful, particularly towards the end of the financial year as patients who wait just less than the current target will breach the target that will operate when the financial year changes.

$I_{[t \in pr]} = 0$ otherwise; and x_{jt} is a set of other time varying covariates which may affect the outcomes. We examine the period from 1st April 1997 to 31st March 2004 and exclude 1999 because it was the year of transition. This includes a pre-policy period (pre-2000) and ends when Scotland begins to implement a waiting times target regime. The data are quarterly. The coefficient of interest is δ.

The outcomes we analyse are various summary statistics of the distribution of the waiting times of patients on the waiting list for elective care; elective admissions; additions to the elective waiting list, length of stay; emergency care admissions; planned admissions; whether patients were re-prioritised; whether patients had complications; and three measures of patient outcomes.

The assumptions required for the difference-in-difference analysis to identify the impact of the policy are that the two countries were subject to the same policies pre-devolution and that the policy change must be exogenous to waiting times. Pre-1999 health care policy was common to both countries. The policy break in 1999 was the result of devolution, which was not related to waiting times for elective care in NHS hospitals. We omit 1999/00 to avoid contaminating pre-policy years with the possible effects of devolution. While the lack of focus on waiting times in Scotland post-devolution may have been due to the perception that waiting times were less important in Scotland than in England, Propper et al. (2008b) show that trends in waiting times were statistically the same in the two countries pre-policy. In our analyses here we test this assumption by fitting a full set of year-country interactions.

2.2 Hospital level analysis

As a robustness check that the difference-in-difference estimates can be attributed to the operation of targets and not some other aspect of devolution or unrelated changes, we examine the responses of hospitals to target pressure. This variation, which is only observed for English hospitals, allows us to test whether any differences between the performance of Scottish and English hospitals at country level are mirrored in differences in performance between English hospitals as a response to differential target pressure.

We define our measure of target pressure as the number of patients waiting in hospital *h* at the end of the previous quarter whose waiting times will exceed the end of the quarter target unless they are treated within the quarter.[9] This is normalised by the total number of patients waiting at the end of the previous quarter. We refer to this as the 'distance from target'.

This measure was initially plausibly exogenous to the hospital. When the target was introduced, the same target was set for all hospitals, thus for some hospitals this meant that targets were easy to achieve while for others the distance to target, and so the pressure on managers, was greater. Further exogenous variation comes from the fact that the targets changed each year. However, behaviour by a manager in response to the target will affect the tightness of the target. For example, if patients with longer waits are treated first in order to meet the target, this will initially reduce the target pressure. But if this treatment is at the expense of treating other patients, this increase in the stock will increase the pressure in the future. In general, responses to target pressure may be quite complex, as the manager can substitute between different activities and also try to influence the inflow of patients. We therefore estimate a reduced form model that allows for dynamics in the outcome measures and for possible measurement error and/or endogeneity.

We estimate various specifications of the general model:

(2) $$o_{ht} = \alpha + \beta d_{h,t-1} + z'_{ht}\theta + \sum_i \gamma_i o_{h,t-i} + \eta_h + v_{ht} \,,$$

where o_{ht} is the outcome measure for hospital h, at time t; $d_{h,t-1}$ is the distance from target for hospital h at the end of the previous quarter, $t-1$; z_{ht} contains controls for the size of the hospital workforce and a full set of time dummies; and the η_h are unobserved hospital effects that are constant over time. The parameter of interest is β, the response in the outcome measure to an increase in target pressure. In the estimation, we weight by list size and report robust standard errors that are clustered at hospital level. The time period is 1st April 2001 to 31st March 2006, which begins from the introduction of the first target and finishes when the 6 months target takes effect.

We estimate equation (2) with both a fixed effects, or within-groups, estimator and using the Arellano-Bond (1991) instrumental variables estimator for dynamic panel data models. This

[9] For example, the number of patients waiting more than 15 months on the 31st December 2000 (the 18 month target maximum first takes effect on the 31st March 2001).

estimation procedure transforms the model into first differences and uses lagged levels of the outcome measures to instrument the endogenous lagged differences. We further treat the hospital workforce variable as endogenous due to measurement error, and the distance from target measure $d_{h,t-1}$ as pre-determined, meaning that $d_{h,t-1}$ may be correlated with $v_{h,t-1}$. In the model for admissions, $d_{h,t-2}$ was found to be correlated with the error in the differenced model and we adjusted the instrument set accordingly. The Generalised Method of Moments (GMM) estimation technique is used to obtain parameter estimates. See the Appendix B for further details.

We use this approach to analyse three possible responses to target pressure. First, we examine admissions. Second, we examine gaming of the waiting list by exploiting data on removals and suspensions that are available for English hospitals only and we examine the proportion of patients added to the waiting list as a planned admission. Third, we examine the impact on patient outcomes.

2.3 Data

The data are from a number of sources, precise details of which are given in Appendix A. A brief overview follows below.

Episode data

Episode data contain information about inpatient and day-case (ambulatory surgical treatment) episodes in the NHS. A record is generated at the end of each episode, when a patient is discharged from care or transferred to the care of another consultant or provider. For periods of care comprising more than one episode (due to transfers), we refer to the first in the sequence as the admission record. Our analysis uses elective, planned[10] and emergency admission records from the Hospital Episode Statistics (HES) database for England and the Scottish Morbidity Record (SMR01) for Scotland. The total annual numbers of these are large (6.4M electives + planned and 4.1M emergencies for England, 0.6M electives + planned and 0.5M emergencies for Scotland in 2003/04). We use random samples of elective and planned admissions (10% for England, 50% for Scotland) and all the data on emergency admissions.

[10] In our analysis we define elective admissions to exclude planned care.

Waiting times for elective admissions are recorded in HES and SMR01 as the difference between the date of decision to admit and the date of admission. We use these to analyse country level waiting times distributions.

We also use the data to calculate total elective admissions, additions to the elective waiting list, and emergency admissions; mean length of stay; planned admissions as a proportion of elective and planned admissions combined; proportion of patients awaiting an 'urgent' elective treatment who are admitted quickly (within 14 days)[11]; proportion of elective admissions with complications[12]; and mortality within 30 days of admission for all patients, emergencies, and AMI emergencies aged 55 and over. All of these are calculated at country level for use in our difference-in-difference analysis. Planned admissions are also calculated and analysed at hospital level.

Census data

Hospital level waiting list data for England are collected on a quarterly basis by the Department of Health and contain information about the stock of patients waiting on a census date (the last day of the month in June, September, December and March) and the flows on and off the list during the quarter. The data are published at provider level and used to monitor performance against NHS waiting time targets.

The data include a breakdown – in three-monthly time bands – of the waits so far experienced by the patients on the waiting list at the census date. We use this to construct the measure of hospital level target pressure as defined in the methodology section above.

The census data also include counts of elective admissions, suspensions and removals during the quarter, which we use in our hospital level analysis. It is not possible to identify which patients are suspended or removed. Since data for Scotland are not collected on an equivalent basis, we are unable to analyse removals and suspensions using our difference-in-difference methodology.

---

[11] In order to define urgent treatment we calculated, for each Healthcare Resource Group (HRG), the proportion of patients admitted within 14 days during the pre-policy period. HRGs for which this proportion exceeded 80 per cent were classified as urgent.
[12] The presence of complications is identified via HRG codes, for those HRGs which differentiate between with and without complications.

A hospital is included in our sample for analysis if it has a waiting list of at least 150. At the end of 2005/06 there were 216 hospitals of which 184 met this criterion.

Other data sources

Data on patient mortality are constructed by matching HES to ONS death records. Our analysis uses the following four outcome measures for the specialties General Surgery and Trauma and Orthopaedics: in-hospital mortality, mortality before or within 28 days of discharge, and mortality within 30 days of selected surgical procedures (elective and non-elective separately). We also look at mortality before or within 28 days of discharge following an AMI and mortality within 30 days of a CABG operation.

In robustness tests of the difference-in-difference analyses we control for health expenditure and need. These data are taken from the Public Expenditure Statistical Analyses (health expenditure) and ONS Population Trends (population estimates and age standardised mortality ratios). Our hospital level analyses use data on the workforce and finances of English hospitals to control for resources. This is collected annually by the Department of Health.

3. Did the targets reduce waiting times?

We begin our analysis with an examination of whether the targets did reduce waiting times. Propper et al. (2008b) showed that the policy reduced the proportion of the persons on the list waiting longer than the targets. Here we examine the whole distribution of waiting times. Figure 1 presents the distribution of realised waiting times for all patients who received elective treatment in the two countries. Patients were classified according to the year in which they were put on the list. The dotted line shows the pre-policy distribution (the same in each (country-specific) panel). The solid line shows the post-policy distribution, where there is one for each year for each country. The vertical dotted lines mark the waiting times targets in operation in England. Comparing pre- and post-policy distributions, it is clear that the effect of the policy in England was to pull the distribution leftwards at the right tail. In contrast in

Scotland the distribution moved rightwards, increasing the number of longer waits and reducing the number that waited below the target set for England.[13]

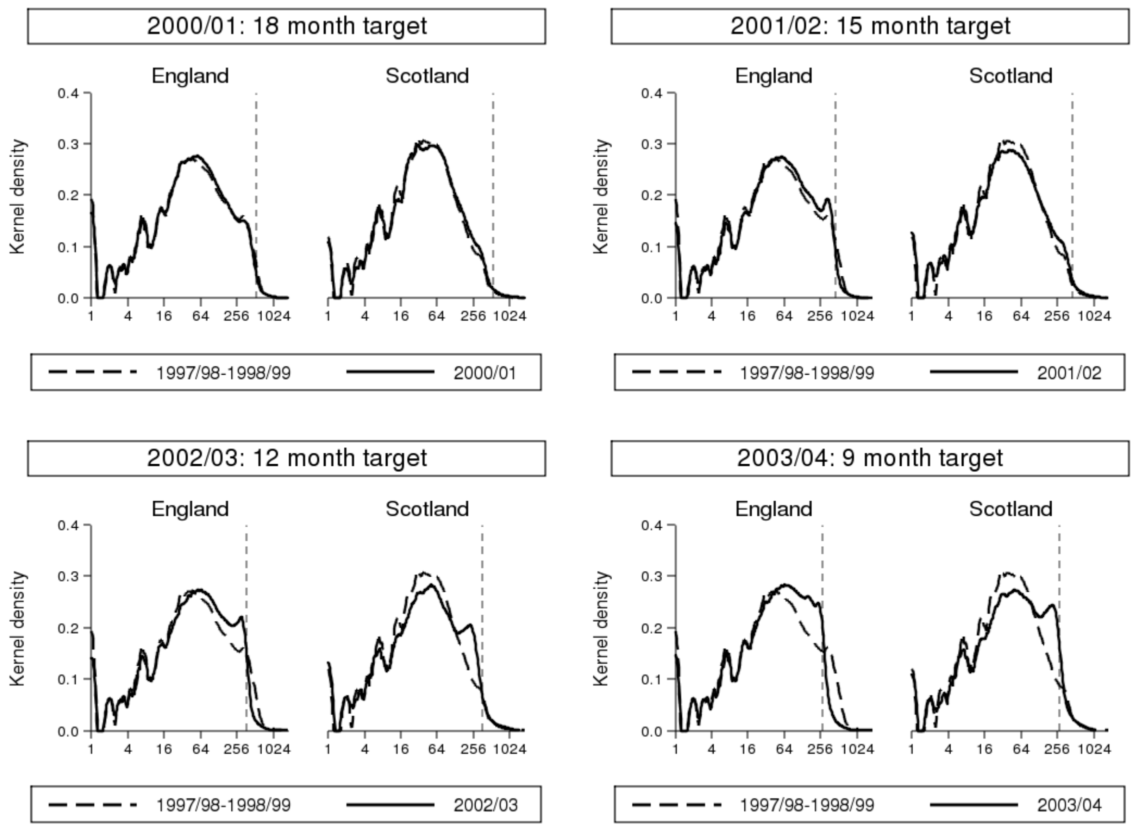
Table 2 presents formal statistical tests of the impact of the target using difference-in-difference estimates of the impact of the policy on various points in the distribution. At the bottom end of the distribution (the 10th and the 25th percentile) the policy appears to have resulted in English waits that are slightly higher. At higher points of the distribution, waits after the policy in England fell significantly. The falls in waiting time are 13 days at the mean and 55 days at the 90th percentile. The results at the mean, the 75th and the 90th percentile are robust to a measure of need, health care expenditure and staffing at the country level.[14]

The model was re-estimated with a full set of time dummies to test the common trends assumption (which is also a test of no anticipation effects). The results (Table A3 in the Appendix) indicate no significant reduction in the proportion waiting prior to the policy introduction except at the very top of the distribution where there is a reduction at the 90th percentile in the two years immediately before the policy in England and one at the 75th percentile in 1999. However, these falls are much smaller than those in any of the post policy years and probably reflect responses to the general concern over very long waits that existed prior to the introduction of the target regime. Falls at other parts of the distribution only started in the first year of the policy.

These analyses indicate that the policy appears to have had the intended effect on waiting times in England. These fell not just to meet the targets, but across the top half of the distribution.

[13] Similar trends have been observed for Wales and Northern Ireland (Alvarez-Rosete et al., 2005), suggesting that this is the trend that would have been observed in England in the absence of the regime. Besley et al. (2008) use a difference-in-difference analysis for England and Wales and also find that targets reduced waiting times.
[14] Introducing these controls slightly increases the d-in-d estimates: the fall in mean waiting time is -16 (se=2.8), the 75th percentile is -21 (se=4.8) and the 90th percentile is -74 (se =13.3) days, respectively.

Figure 1. Distribution of waiting times



Notes: horizontal-axis on the log scale. English data from HES (10% sample of elective spells by financial year of addition to list, 1997/98 to 2003/04). Scottish data from SMR01 (50% sample of elective spells by financial year of addition to list, 1997/98 to 2003/04). Vertical line is positioned at the English end-of-year target maximum wait.

Table 2. Impact of policy on waiting times (days)

	10th percentile	25th percentile	Median	75th percentile	90th percentile	Mean
Constant	4.85**	13.31**	34.48**	79.82**	168.10**	67.79**
	(0.19)	(0.26)	(0.69)	(2.53)	(6.62)	(1.39)
Year	0.06	0.14	1.30**	3.61**	-3.45	-0.95
	(0.08)	(0.11)	(0.29)	(1.05)	(2.76)	(0.58)
England	-1.00**	0.50	7.25**	35.13**	105.13**	26.29**
	(0.26)	(0.36)	(0.95)	(3.50)	(9.15)	(1.93)
Policy	-0.66	-0.61	-2.24	0.42	48.19**	12.01**
	(0.38)	(0.53)	(1.41)	(5.19)	(13.59)	(2.86)
Policy*England	**0.94***	**1.13***	**1.00**	**-11.44***	**-55.44***	**-12.66***
	(0.31)	**(0.44)**	**(1.16)**	**(4.28)**	**(11.21)**	**(2.36)**
Number of obs	48	48	48	48	48	48

Notes: English data from HES (10% sample of elective spells by quarter of addition to list, 1997/98 to 2003/04). Scottish data from SMR01 (50% sample of elective spells by quarter of addition to list, 1997/98 to 2003/04). Pre-policy period is 1997/98 to 1998/99; post-policy period is 2000/01 to 2003/04. 1999/00 omitted. Observations are at quarterly country level. Standard errors in parentheses. Significance levels: * 5% ** 1%

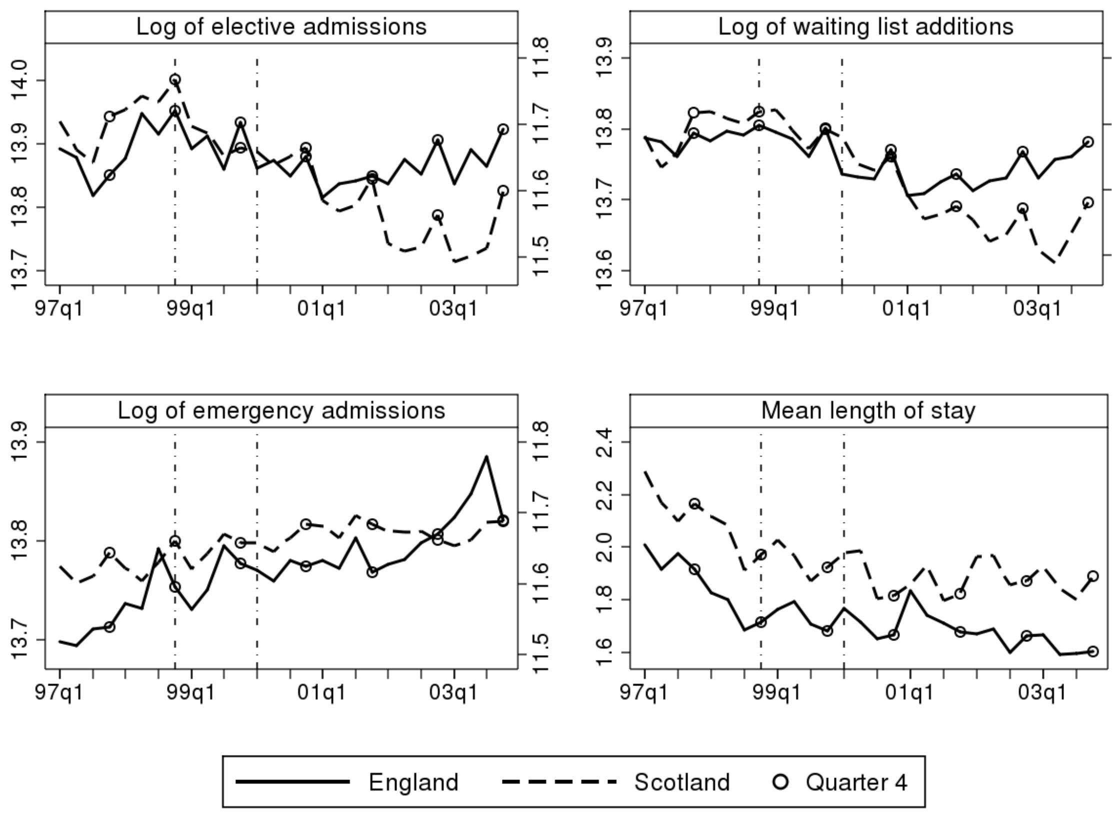
4. Reallocation of activity, gaming and patient outcomes

4.1 Effort on monitored and unmonitored tasks

Waiting times fell as a result of the policy. To achieve this, hospitals may have diverted effort towards getting elective patients treated, but simultaneously added fewer patients onto the elective waiting list and so avoided the necessity of increasing effort in the future. They could have also reduced effort on the treatment of elective patients by discharging them sooner or reduced effort on non-elective patients.

We begin by examining whether hospitals increased the number of electives treated. The top row of Figure 2 presents the country level trends in hospital activity for the financial years 1997 to 2003. In both countries, elective admissions and additions to the list appear to be increasing before the policy. After the policy, both fell in Scotland, whilst in England the trends level out, with admissions and additions both increasing towards the end of the period.

Figure 2. Country level activity, 1997/98 to 2003/04



Notes: English data from HES, 1997/98 to 2003/04. Scottish data from SMR01, 1997/98 to 2003/04. Excepting the figure for length of stay, the vertical-axis for Scotland is on the right hand side. Vertical lines (from left to right) mark the last quarter of the pre-policy period and the first quarter of the post-policy period.

Table 3 presents the difference-in-difference estimates. The first two columns show the policy resulted in an 11% increase in elective admissions and an 8% increase in additions to the waiting lists. Activity in elective care therefore increased in response to targets: English hospitals both admitted more patients and treated them faster.

We next examine diversion of activity: did hospitals achieve a greater volume of elective care by reducing effort on the treatment of patients once in hospital or by reducing the number of non-electives patients they treated? As a measure of the first we examine lengths of stay for elective patients; as a measure of the second, we examine number of emergency admissions. Figure 2 shows little support for a reduction in effort on these two activities. The number of emergency admissions increases steadily in both countries. The mean length of stay decreased in both countries before 1999/00, after which it appears fairly level. Columns 3 and 4 of Table 3 test for country differences. This confirms the graphical picture: there is no impact of the policy on either length of stay or emergency admissions.

4.2 Did managers respond by gaming waiting lists?

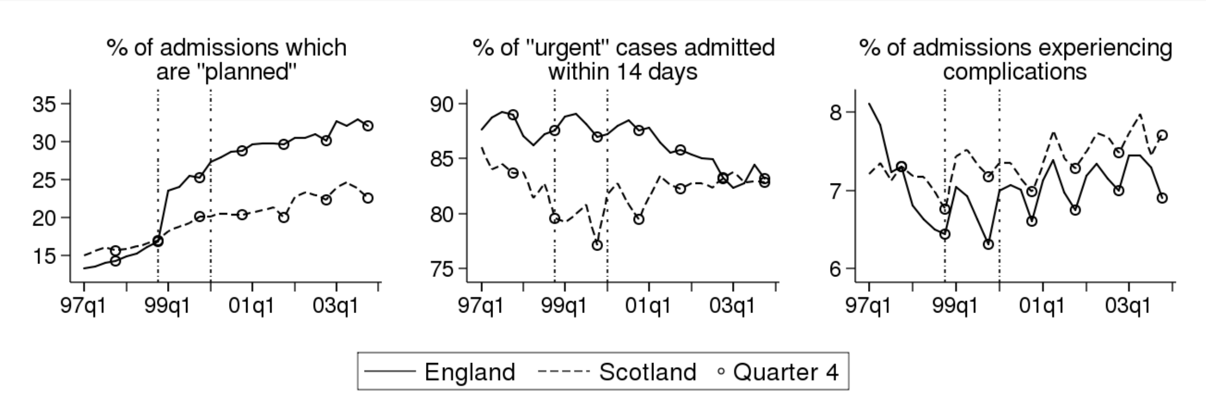
Reclassification

As noted above, patients whose treatment was classified as planned did not count towards waiting times targets. The list subject to target could also be reduced by suspending or removing patients. The use of suspensions and removals cannot be examined in a difference-in-difference framework because these data were not collected in a comparable fashion in England and Scotland but it is possible to examine differences pre- and post-policy in the volume of patients classified as planned.

The first panel in Figure 3 shows that the volume of planned activity rose in England compared to Scotland. The difference-in-difference estimates, in column 5 of Table 3, show that there is a significant increase in planned activity in England post-policy.[15] However, examination of the figure shows that the initial rise in planned in England actually occurs pre-policy and so appears to be a continuation of a change that first occurred before the policy rather than a response to target pressure.

[15] Note the difference-in-difference estimates omit the first year of this rise, 1999, because it is a transition year.

Figure 3. Country trends in planned admissions, urgent admissions and admissions with complications, 1997-2004



Notes: English data from HES (10% sample of elective spells by quarter of admission, 1997/98 to 2003/04). Scottish data from SMR01 (50% sample of elective spells by quarter of admission, 1997/98 to 2003/04). Vertical lines (from left to right) mark the last quarter of the pre-policy period and the first quarter of the post-policy period.

Reprioritisation

If managers responded to pressure by changing the ordering of treatment of patients, replacing treatment by medical severity with treatment by length of time on the lists, patients who pre-policy waited shorter times would wait longer post policy. The difference-in-difference estimates for waiting time do show an increase in waiting times at the lower end of the waiting time distribution. To examine this further we look at whether urgent cases had longer waits post-policy and whether more patients were admitted with complications post-policy, the latter possibly indicating that sicker patients had to wait longer and so developed more complications whilst waiting.

The second panel of Figure 3 shows the proportion of urgent cases (as defined in the data section) admitted within two weeks in the two countries. It is clear from the figure that Scotland admitted a lower proportion of these cases within a fortnight. It is also clear that the countries diverge, with the number in England trending downwards over the period while the Scottish proportions fall and then rise. A simple difference-in-difference estimator would show that there is a fall in England relative to Scotland of around 2 percent. However, the figure also shows the two countries diverge before devolution. Estimation of a difference-in-difference model allowing for full country-year interactions shows the common trend pre-policy assumption is not supported by the data (results from authors), so using our country

Table 3. Difference-in-difference estimates

	Activity (log transformed):			(4) Mean LOS	(5) % Planned admissions	(6) % With complication	Within 30 day mortality rate:		
	(1) Elective admissions	(2) List additions	(3) Emergency admissions				(7) All admissions	(8) AMI admissions	(9) Emergency admissions
Constant	11.72**	11.70**	11.62**	2.12**	15.28**	-3.23	2.13**	21.39**	4.67**
	(0.02)	(0.01)	(0.01)	(0.03)	(0.21)	(15.68)	(0.05)	(0.26)	(0.10)
Financial year	-0.01	-0.01*	0.01**	-0.03**	1.31**	0.06	-0.01	-0.54**	-0.04
	(0.01)	(0.01)	(0.00)	(0.01)	(0.09)	(0.11)	(0.02)	(0.11)	(0.04)
England	2.18**	2.10**	2.10**	-0.24**	-1.21**	0.41	-0.38**	-3.58**	-0.85**
	(0.02)	(0.02)	(0.01)	(0.04)	(0.29)	(0.46)	(0.07)	(0.36)	(0.14)
Policy in operation	-0.09**	-0.07*	-0.01	-0.08	0.73	0.92**	-0.05	0.47	-0.22
	(0.03)	(0.03)	(0.02)	(0.06)	(0.43)	(0.33)	(0.10)	(0.54)	(0.21)
Policy*England	**0.11****	**0.07****	**0.02**	**0.04**	**9.48****	**-0.54***	**-0.33****	**-2.10****	**-0.27**
	(0.03)	**(0.02)**	**(0.01)**	**(0.05)**	**(0.35)**	**(0.20)**	**(0.08)**	**(0.44)**	**(0.17)**
Number of obs	48	48	48	48	48	48	48	48	48

Notes: English data from HES (10% sample of elective spells, 1997/98 to 2003/04; all emergency admissions, 1997/98 to 2003/04). Scottish data from SMR01 (50% sample of elective spells, 1997/98 to 2003/04; all emergency admissions, 1997/98 to 2003/04). Pre-policy period is 1997/98 to 1998/99; post-policy period is 2000/01 to 2003/04. 1999/00 omitted. Observations are at quarterly country level. All variables except for 30 day mortality rate post admission for emergency AMI pertain to the whole hospital. For analysis of complications we include case-mix controls (age breakdown of admitted patients in bands (0-15, 16-24, 25-34, 35-49, 50-69, >=70)). Standard errors in parentheses. Significance levels: * 5% ** 1%.

difference-in-difference design we cannot conclude that the policy increased length of waits for very urgent cases.

To look further at reshuffling, we examine the complication rate. Figure 3 shows a large fall in complications in England pre-policy but a very similar pattern post policy in the two countries. Column (6) of Table 3 presents the difference-in-difference estimate. As expected from the figure the estimate is negative, indicating fewer complications in England post-policy. But the difference is not significant and appears to be driven by pre- rather than post-policy divergence in trends.

4.3 The effect of the policy on quality

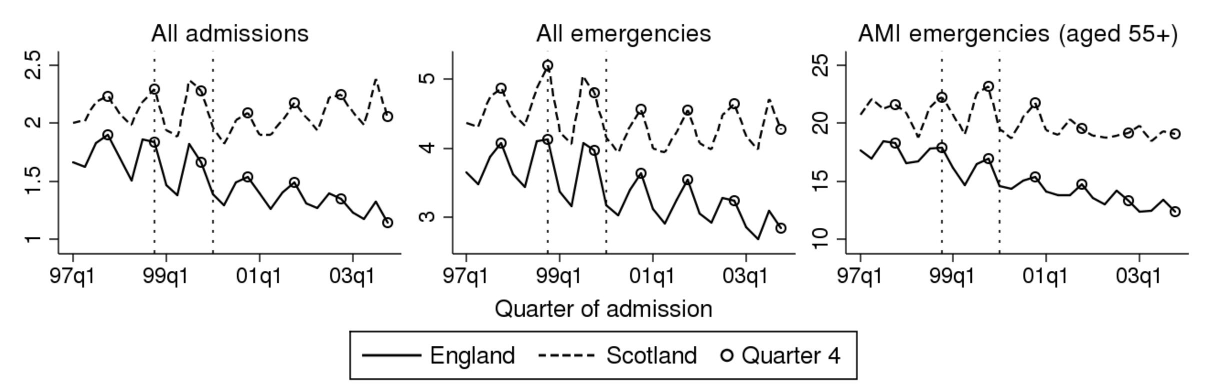
We examine three measures of mortality as a proxy for quality: the within 30 day mortality rate for all admissions, all emergency admissions and emergency admissions for acute myocardial infarction (AMI). Figure 4 shows the time trends in the two countries by quarter. While all the mortality series show strong seasonal effects, the trends in England appears to be decreasing faster, most visibly for mortality for all admissions. Thus England appears to have better, and not worse, outcomes post- policy. The difference-in-difference estimates in columns (7)-(9) of Table 3 confirm this. Post-policy the within 30-day death rate was 0.33 percentage points lower in England and the AMI death rate was 2.1 percentage points lower. These are quite large falls relative to the respective means of 1.7 and 15.8. There was no significant change in the death rate post emergency admission.

5. Hospital level analyses

The country level analyses show that the policy decreased waiting times without clear evidence of diversion of effort from other activities or a fall in patient outcomes. It is possible that these results are not due to pressure from the waiting times targets per se but derive from some other aspect of the policy regime in England.[16] One possibility is that it was simply the greater focus on the behaviour of the individual hospital in England compared to Scotland. If this is correct, then we would expect no relationship between the pressure that the hospital was under from the waiting time target and the behaviour of the hospital. We therefore use the

[16] One candidate we have already examined, and found not to be driving the results, was the growth of health care resources at country level.

Figure 4. Percentage dying within 30 days of treatment by country, quarterly



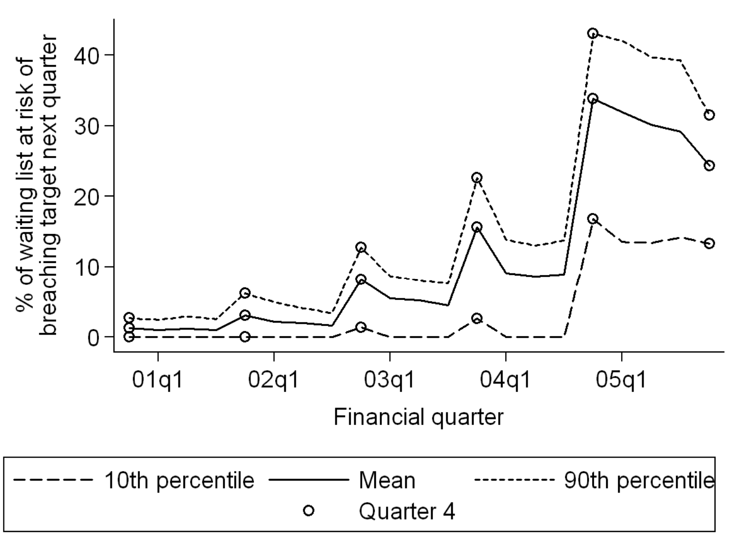
Notes: English data from HES (10% sample of elective spells by quarter of admission, 1997/98 to 2003/04, all emergency spells by quarter of admission, 1997/98 to 2003/04). Scottish data from SMR01 (50% sample of elective spells by quarter of admission, 1997/98 to 2003/04, all emergency spells by quarter of admission, 1997/98 to 2003/04). Vertical lines (from left to right) mark the last quarter of the pre-policy period and the first quarter of the post-policy period.

between and within hospital variation in distance from targets in English hospitals to directly examine the association between the probability that a hospital will breach its target at the end of the current quarter and outcomes of interest. We examine admissions and then look more closely at gaming, examining first the unexpected result for planned admissions and then looking at removals and suspensions. Finally, we examine various mortality measures.

5.1 Pressure from the targets

Figure 5 shows the distribution of the pressure measure across hospitals. It is clear that targets got tighter each year. At the beginning of 2001, the mean proportion of patients at risk of breaching the target was very small for all hospitals, reflecting the fact that there were few individuals who actually waited over 18 months. However, by the end of the period we analyse, when the targets had tightened to 6 months, the mean proportion was over 20%. There are sharp rises at the end of each financial year followed by falls within the year. While we would expect an increase in the proportion of patients at risk of breaching the target as the targets tightened each year, the sharp rise at the end of each year indicates that hospitals did not change activity sufficiently in advance of a change in targets to keep the proportion of patients at risk constant. There is also substantial variation between hospitals in the pressure that they faced. Over 10% of hospitals faced no pressure throughout the period with the exception of the final year and the fourth quarters when the targets tightened. In contrast, by the end of the period 10% of hospitals found themselves with two in every five waiting patients threatening to breach the target if not admitted within the next quarter.

Figure 5. Distribution of the pressure measure across hospitals



Notes: England only; data from Department of Health waiting times/list statistics, quarter 4 of 2000/01 to quarter 4 of 2005/06.

5.2 Admissions

We use equation (2) to test whether the increase in elective admissions seen in the country level analysis is associated with distance from target. Table 4 presents the results. We show first the static fixed effects, or within groups (hospital) estimates (column 1) and test whether this model is misspecified in terms of dynamics by examining the serial correlation in the residuals. We then allow for dynamics using the within groups estimator (column 2), where the number of lags of the dependent variable is determined by the tests for serial correlation. We then present our preferred Arellano-Bond GMM estimates, which allow for dynamics and endogeneity, in column 3.

Column 1 shows clear evidence of serial correlation in the residuals when dynamics are omitted. Column 2 allows for lagged and twice-lagged admissions. With both lags the hypothesis of no serial correlation in the residuals cannot be rejected. The estimate on the pressure variable is significant. The coefficient estimate shows a 1 point increase in the pressure variable (which runs from 0 to 100) leads to a 0.2% increase in elective admissions. The GMM estimates pass the specification tests. The dynamics are similar to the fixed effects estimates of column (2), but the estimates of the effect of distance from target are somewhat larger for the GMM estimates. A 1 percentage point increase in target pressure results in just under a 1% increase in admissions.

Table 4. Estimated effect of target pressure on admissions and planned additions

	Admissions			**Planned Additions**		
	Fixed Effects		GMM	Fixed Effects		GMM
Target pressure_1	0.001 (0.001)	0.002** (0.001)	0.008** (0.002)	-0.005 (0.008)	-0.003 (0.003)	-0.014 (0.011)
DepVar_1		0.481** (0.056)	0.406** (0.098)		0.658** (0.041)	0.665** (0.160)
DepVar_2		0.148** (0.031)	0.107** (0.032)		0.082** (0.031)	0.044 (0.034)
AR1 (p)	0.00	0.14	0.00	0.00	0.36	0.00
AR2 (p)	0.00	0.16	0.43	0.00	0.42	0.39
Hansen (p)			0.19			0.52
Dof			5			5
N	217	213	189	148	148	147
NT	3572	3184	2967	2594	2288	2138

Notes: Dependent variables are log admissions and log planned additions. Estimation routine used is xtabond2 in Stata, Roodman (2006). All regressions weighted by list size. Robust standard errors, clustered at hospital level, in parentheses. Observations are quarterly and cover the sub-sample of hospitals for which all relevant data are available. DepVar_j is the jth lag of the dependent variable. Size of workforce variable included in the model. AR are tests for serial correlation in the residuals of the models. For the FE models these are directly on v_{ht}. The null is that of no serial correlation between v_{ht} and $v_{h,t-j}$. For the GMM results these are tests for serial correlation in the differenced errors, Δv_{ht}. Rejection of AR2 indicates serial correlation in v_{ht} in that case. AR tests for fixed effects models are based on unweighted estimation results. GMM estimates are first-differenced one-step GMM, using instruments dated t-2,...,t-4, collapsed. For admissions, the lagged levels of the target pressure measure dated t-3,…,t-5 are used as instruments. Hansen is a test for instrument validity. Dof is the number of overidentifying instruments. Significance levels: * 5% ** 1%. See further details in Appendix B.

5.3 Gaming

We first examine whether the increase in planned additions seen in the country level analysis appears to be related to responses to target pressure. Table 4 shows clear evidence of dynamics, but no evidence that planned additions are associated with target pressure.[17] We also examine the impact of the policy on the shorter-term strategies of increasing removals and suspensions. Table 5 presents the estimates. The within groups models with no dynamics are misspecified but, even in this model, there is an association of removal activity with the target pressure. The within groups estimators allowing for dynamics are significant and

[17] The results are very similar using planned additions as a proportion of all additions (available from the authors).

positive and show no evidence of dynamic misspecification. The preferred GMM estimator increases the point estimates for the target pressure variable quite considerably compared to the within group estimates. A 1 percentage point increase in the pressure variable increases suspensions and removals by about 2 and 2.5 percent respectively.

Table 5. Estimated effect of target pressure on suspensions and removals

	Suspensions			**Removals**		
	Fixed Effects		GMM	Fixed Effects		GMM
Target pressure_1	0.008 (0.005)	0.006** (0.002)	0.019** (0.007)	0.012** (0.004)	0.008** (0.002)	0.027** (0.006)
DepVar_1		0.813** (0.027)	0.732** (0.180)		0.466** (0.071)	0.549** (0.067)
DepVar_2					0.134* (0.061)	0.190* (0.079)
AR1 (p)	0.00	0.37	0.00	0.00	0.48	0.00
AR2 (p)	0.00	0.77	0.62	0.01	0.25	0.28
Hansen (p)			0.27			0.45
Dof			6			5
N	213	212	187	217	213	189
NT	3468	3266	3043	3570	3180	2963

Notes: Dependent variables are log number of suspension and log number of removals from the list. See further notes of Table 4.

We also examine the issue of re-prioritisation of patients by looking at the proportion of urgent cases admitted within 14 days and the proportion of cases with complications as a function of target pressure. We found no effect of distance from target on either outcome (results available from authors).

These analyses suggest that target pressure did lead managers to admit more patients; it also led to some gaming, in terms of increasing the number of suspensions and removals. But they suggest that the somewhat puzzling finding of an increase in planned cases in the difference-in-difference results was not due to the waiting list policy.

5.4 Magnitudes of the impact of targets

These estimates can be used to provide an estimate of the impact on behaviour from target pressure. The estimates show a one percentage point increase in the pressure measure resulted in just under a 1% increase in admissions, a 2% increase in removals and a 2.5% increase in suspensions. Evaluated at the relevant mean for all hospitals, this equals approximately 40 extra admissions, 14 more removals and 8 more suspensions per quarter.

A one percentage point increase in the pressure measure is small compared to the change in the measure over time and is also small compared to the gap between those hospitals most at risk and those least at risk (see Figure 2). After removing the time trend in the pressure measure, the between hospital 90:10 percentile gap in the target pressure distribution is approximately 10 percentage points. Using our estimates, a hospital at the 90th point in the distribution of target pressure compared to one at the 10th point in this distribution would admit 400 more patients per quarter. This is a reasonably large number of extra admissions, though only about 5% of the 90:10 gap in admissions. For suspensions, a hospital at the 90th point in the target pressure distribution would have 70 more suspensions, which is about 10% of the 90:10 gap in suspensions. For removals, a hospital at the 90th percentile of the target pressure distribution would have 180 more removals, which is just over 13% of the 90:10 gap in removals.[18]

These estimates can be compared to those from the country analysis to check that the target pressure measure is in the same ball park as the difference-in-difference estimates. The difference-in-difference results showed a increase in admissions between England and Scotland post policy of around 11% per quarter. Estimated at the mean number of admissions per English hospital of 4000 this equates to 440 more admissions per quarter. This is of a similar magnitude to the difference in admissions between a hospital in the top and one in the bottom decile of the target pressure distribution.

5.5 The effect on patient outcomes

The country level analyses showed that patient outcomes did not deteriorate post policy in England. In fact, they improved on two of the three measures. We examine the robustness of these findings by estimating the association of mortality with target pressure. It is unlikely that target pressure is affected by mortality of patients and therefore we assume the target pressure variable is exogenous to mortality. We therefore estimate fixed effects models for a range of mortality measures controlling for mortality specific case-mix. Table 6 presents the results. None of the coefficients are either large or statistically significantly different from

[18] The 90:10 gap in admissions is 7800, in suspensions is 750 and in removals is 1360.

zero. We conclude there is little evidence that quality, as measured by mortality rates, fell as a result of targets.

Table 6. Impact of distance from target on patient outcomes (per 10,000)

	In-hospital mortality	Mortality within 28 days of discharge		Mortality within 30 days of surgery (selected procedures)		
		All admissions	AMI aged 55+	Elective	Non-elective	CABG
Specialty level						
General surgery	0.19	0.22		-0.14	0.20	
	(0.15)	(0.19)		(0.11)	(0.55)	
Trauma and orthopaedics	-0.18	-0.25		-0.02	-0.47	
	(0.15)	(0.20)		(0.05)	(0.38)	
Hospital level			5.68			0.85
			(4.10)			(1.21)

Notes: Data from HES 2001/02 to 2005/06, linked to ONS mortality records. Analysis at individual episode level. Analyses of 'in hospital' and 'within 28 days of discharge' mortality use 50% samples of the full dataset. Robust standard errors in parentheses. Regressions contain hospital fixed effects. For specialty level outcomes, we control for quarter of admission, age and gender of patient, HRG dummies, nurses as % of staff, doctors as % of staff, and log of total admissions. For hospital level outcomes, we control for quarter of admission, age and gender of patient, non-elective admission indicator (CABG only), indicator for 'with complications' (AMI only), nurses as % of staff, doctors as % of staff, and log of total admissions. Significance levels: * 5% ** 1%.

6. Conclusions

This paper provides evidence that, contrary to popular views, a policy of targets for waiting lists in the English NHS appears to have achieved its objectives. The length of time patients waited fell and admissions for elective care rose. This fall in waiting times was achieved without many of the gaming activities that had been forecast. The waiting times distribution did not stack up at the maximum waiting point, the order in which patients were treated from the list did not appear to change, the proportion of urgent cases treated did not fall and there is no evidence of a decrease in several measures of quality of care as a result of the policy.

However, there is also evidence that the policy did lead to waiting list manipulation: the number of suspensions and removals increased as a result of the policy. On the measures of severity and patient care that we have available, we find no evidence that the severity of patients who were admitted for treatment changed or that the quality of care of those treated in hospital fell. So we conclude that this was probably not gaming, in the sense that it was associated with welfare losses, but was either simply better management of lists or occurred on too small a scale to affect overall patient outcomes. However, we note that the quality measures we use are quite crude and pertain to treated patients. It is possible that there are patients who were worse off because of this policy: for example, suspended patients received treatment later than they otherwise would have done and removed patients possibly received no hospital treatment at all. But our measures do not pick up their possibly poorer outcomes.

Our findings raise the issue of why the policy appears to have met its aim with no evidence of negative side effects. One possibility is the one we just mention: our quality measures are too broad and hospital based to capture any fall in outcomes. But given the positive outcomes that we do find, we suggest three other possible and non-exclusive reasons. First, because long waiting lists were seen as a problem by most of society, the targets may have acted as a mission for NHS employees, inducing additional effort at no cost. Second, the fact that the targets were announced in advance and were escalating may have meant that production was reorganised on a long term basis, so increasing productivity over the long term, rather than simply resulting in short term fixes to a once-off policy. Finally, the policy was accompanied by extra resources. While we have established that the results are robust to controls for resources, it is probably easier to engage in service reorganisation in a time of generous resources.

References

Alvarez-Rosete, A., G. Bevan, N. Mays and J. Dixon (2005), Effects of Diverging Policy Across the NHS, *British Medical Journal* 331, 946-950.

Arellano, M. and S. Bond (1991), Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies* 58, 277-298.

Baker, G. (1992), Incentive Contracts and Performance Measurement, *Journal of Political Economy* 100, 598-614.

Besley, T., G. Bevan and K. Burchardi (2008), Accountability and Incentives: The Impacts of Different Regimes on Hospital Waiting Times in England and Wales, mimeo, London School of Economics.

Besley, T., J. Hall and I. Preston (1999), The Demand for Private Health Insurance: do Waiting Lists Matter?, *Journal of Public Economics* 72, 155-181.

Besley, T. and M. Ghatak (2003), Incentives, Choice, and Accountability in the Provision of Public Services, *Oxford Review of Economic Policy* 19, 235-249.

Besley, T. and M. Ghatak (2005), Competition and Incentives with Motivated Agents, *American Economic Review* 95, 616-636.

Bevan, G. and C. Hood (2006a), What's Measured is What Matters: Targets and Gaming in the English Public Health Care System, *Public Administration* 84, 517-538.

Bevan, G. and C. Hood (2006b), Have Targets Improved Performance in the English NHS?, *British Medical Journal* 332, 419-22.

Courty, P. and G. Marschke (1997), Measuring Government Performance: Lessons from a Federal Job-Training Program, *American Economic Review* 87, 383-388.

Department of Health (1995), *Patient's Charter*, London, Department of Health.

Department of Health (2000), *The NHS Plan: A Plan for Investment, A Plan for Reform*, Cm 4818-I, Stationery Office, London.

Dewatripont, M., I. Jewitt and J. Tirole (1999), The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies, *Review of Economic Studies*. 66, 199-217.

Dusheiko, M., M. Goddard, H. Gravelle and R. Jacobs (2008), Explaining Trends in Concentration of Healthcare Commissioning in the English NHS, *Health Economics* 17, 907-926.

EWHC 2787 (QB) (2006), Henry vs British Broadcasting Corporation: Case No. HQ04X02769, Royal Courts of Justice, London.

Friedman, J. and S. Kelman (2007), Effort as Investment: Analyzing the Response to Incentives, Working Paper Series rwp07-024, Harvard University, John F. Kennedy School of Government.

Gaynor, M. (2006), What do we Know about Competition and Quality in Health Care Markets?, CMPO Discussion Paper 06/151, Department of Economics, University of Bristol.

Goddard M., R. Mannion and B. Ferguson (1997), Contracting in the UK NHS: Purpose, Process and Policy, Working Paper 156chedp, Centre for Health Economics, University of York.

Heckman, J., C. Heinrich and J. Smith (1997), Assessing the Performance of Performance Standards in Public Bureaucracies, *American Economic Review* 87, 389-395.

Holmstrom, B. and P. Milgrom (1991), Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design, *Journal of Law, Economics and Organization* 7, 24-52.

Hood, C. and G. Bevan (2005), Governance by Targets and Terror: Synecdoche, *Westminster Economics Forum* 15.

Kelman, S. and J. Friedman (2007), Performance Improvement and Performance Dysfunction: An Empirical Examination of Impacts of the Emergency Room Wait-Time Target in the English National Health Service, Working Paper Series rwp07-034, Harvard University, John F. Kennedy School of Government.

Kessler, D. and M. McClellan (2000), Is Hospital Competition Socially Wasteful?, *The Quarterly Journal of Economics* 115, 577-615.

Laing, W. (2007), *Laings Review of Private Health Care*, Laing and Buisson, London.

Lavy, V. (2008), Performance Pay and Teachers' Effort, Productivity and Grading Ethics, *American Economic Review*, forthcoming.

National Audit Office (2001), *Inpatient and Outpatient Waiting in the NHS*, Report by the Comptroller and Auditor General HC211, London.

Osborne, D. and T. Gaebler (1992), *Reinventing Government*, Addison-Wesley.

Prendergast, C. (2003), The Limits of Bureaucratic Efficiency, *Journal of Political Economy* 111, 929-958.

Propper, C., S. Burgess and D. Gossage (2008a), Competition and Quality: Evidence from the NHS Internal Market 1991-9, *Economic Journal* 118, 138-170.

Propper, C., M. Sutton, C. Whitnall and F. Windmeijer (2008b), Did 'Targets and Terror' Reduce Waiting Times in England for Hospital Care?, *The B.E. Journal of Economic Analysis & Policy* (2008), 2 (Contributions), Article 5.

Propper, C. and D. Wilson (2003), The Use and Usefulness of Performance Measures, *Oxford Review of Economic Policy* 19, 250-267.

Roodman, D. (2006), How to Do xtabond2: An Introduction to "Difference" and "System" GMM in Stata, Working Paper 103, Center for Global Development, Washington.

Scottish Executive (2000), *Our National Health: A Plan for Action, a Plan for Change*, Scottish Executive, Edinburgh.

Scottish Executive (2003), *Partnership for Care: Scotland's Health White Paper*, Scottish Executive, Edinburgh.

Smith, P. (1995), On the Unintended Consequences of Publishing Performance Data in the Public Sector, *International Journal of Public Administration* 18, 277-310.

Appendix A. Data Sources

Table A1: Data sources for country level outcomes, 1997/98 to 2003/04

Variable	Notes	Source
% of list waiting x months or more on census date (official)	• At end of financial quarter	Department of Health: Provider based waiting times/list statistics. [i] ISD Scotland: SMR3 Inpatient/Day case waiting list census. [ii]
% of list waiting x months or more on census date (reconstructed)	• At end of financial quarter • Uses discharge records up to 2005/06 to avoid end-of-period truncation	Department of Health: Hospital Episode Statistics (HES) - 10% of elective episodes. ISD Scotland: SMR01 General/Acute Inpatient and Day Case - 50% of elective episodes.
Elective admissions: - Volume - Waiting duration distribution - % with complications - % "planned" - Mean length of stay - "Urgent" HRGs	• During financial quarter • Planned admissions are expressed as a percentage of total planned and non-planned elective admissions • All other measures exclude planned admissions	Department of Health: Hospital Episode Statistics (HES) - 10% of elective episodes. ISD Scotland: SMR01 General/Acute Inpatient and Day Case - 50% of elective episodes.
Volume of emergency admissions	• During financial quarter	Department of Health: Hospital Episode Statistics (HES) - all emergency episodes. ISD Scotland: SMR01 General/Acute Inpatient and Day Case - all emergency episodes.
Volume of list additions	• During financial quarter • Uses discharge records up to 2005/06 to avoid end-of-period truncation	Department of Health: Hospital Episode Statistics (HES) - 10% of elective episodes. ISD Scotland: SMR01 General/Acute Inpatient and Day Case - 50% of elective episodes.
Age standardised mortality ratio	• Calendar year • 1997/98 and 1998/99 figures interpolated	ONS Population Trends 125, Table 2.2. [iii]

Health expenditure	• During financial year • Figures for 1997/98 and 1998/99 re-adjusted according to PESA 2005 trend	PESA 2003 table 8.3a, PESA 2004 8.5a and PESA 2005 tables 8.5a, 8.6a, …, 8.9a. [iv]
Workforce whole time equivalent	• At 30th September	NHS hospital and community health services non-medical staff in England: 1994-2004 (Table 1A). [v] NHSScotland Workforce statistics: Table A1. [vi]
Population	• Mid-year estimates	ONS Population Trends 125, Table 1.2. [iv]

Table A2: Data for provider level outcomes, England only, 2000/01 (q4) to 2004/05 (q3) unless otherwise stated

Variable	Notes	Source
Volume of elective admissions	• During financial quarter	Department of Health: Provider based elective admission events. [i]
Volume of emergency admissions	• During financial quarter • 2000/01 (q4) to 2003/04 (q4)	Department of Health: Hospital Episode Statistics (HES) - all emergency episodes.
"Planned" additions as a % of total list additions	• During financial quarter • Uses discharge records up to 2005/06 to avoid end-of-period truncation	Department of Health: Hospital Episode Statistics (HES) - 10% of elective episodes.
List removals	• At end of financial quarter	Department of Health: Provider based elective admission events. [i]
List suspensions	• At end of financial quarter	Department of Health: Provider based patients who have deferred admission. [i]
% of "urgent" HRGs admitted within 14 days	• During financial quarter	Department of Health: Hospital Episode Statistics (HES) - 10% of elective episodes.
% of elective admissions with complications	• During financial quarter • Identified via HRG v3.5 title	Department of Health: Hospital Episode Statistics (HES) - 10% of elective episodes.

Mortality in-hospital and after discharge	• During financial quarter • Defined according to NHS mortality rate PI's for 2002/03 [vii]	HES data linked with ONS death records.
Emergency readmission after treatment for a fractured hip	• During financial quarter • Defined according to NHS re-admission rate PI's for 2002/03 [viii]	Department of Health: Hospital Episode Statistics (HES).
"Distance from target": % of specialty list at risk of breaching target if untreated by next census date	• At start of quarter (as proxied by end of previous quarter)	Department of Health: Provider based waiting times/list statistics. [i]
Workforce: - Whole time equivalent. - Doctors and qualified nurses as % of total WTE.	• At 30th September	Department of Health: Non-medical staff groups by NHS organisation, table 8. [ix] Department of Health: HCHS medical and dental staff, table 3.1. [ix]
List size	• At start of quarter (as proxied by end of previous quarter)	Department of Health: Provider based waiting times/list statistics. [i]

i http://www.performance.doh.gov.uk/waitingtimes/index.htm
ii Obtained on request from ISD
iii http://www.statistics.gov.uk/downloads/theme_population/PT125_main_part3.pdf
iv http://www.hm-treasury.gov.uk/economic_data_and_tools/finance_spending_statistics/pes_publications/pespub_index.cfm
v http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsStatistics/DH_4106723
vi http://www.isdscotland.org/isd/workforce-statistics.jsp?pContentID=1347&p_applic=CCC&p_service=Content.show&
vii http://www.performance.doh.gov.uk/nhsperformanceindicators/2002/trdca_t.doc
viii http://www.performance.doh.gov.uk/performanceratings/2002/s_133.doc
ix http://www.dh.gov.uk/en/Publicationsandstatistics/Statistics/StatisticalWorkAreas/Statisticalworkforce/DH_4087066

Table A3: Difference-in-difference estimates of impact of policy on waiting times (days) including pre- and post-policy year dummies

All specialties	10th percentile	25th percentile	Median	75th percentile	90th percentile	Mean
Constant	5.00**	13.75**	36.25**	84.50**	176.50**	69.91**
	(0.20)	(0.28)	(0.80)	(2.52)	(4.13)	(1.02)
Country = England	-1.25**	-0.25	6.00**	36.75**	113.50**	26.61**
	(0.28)	(0.40)	(1.13)	(3.56)	(5.84)	(1.44)
Year = 1998/99	-0.25	-0.75	-2.25	-5.75	-20.25**	-5.19**
	(0.28)	(0.40)	(1.13)	(3.56)	(5.84)	(1.44)
Year = 1999/00	0.00	0.00	-1.25	-2.75	-4.75	-1.18
	(0.28)	(0.40)	(1.13)	(3.56)	(5.84)	(1.44)
Year = 2000/01	0.00	0.25	1.00	4.25	9.00	3.48*
	(0.28)	(0.40)	(1.13)	(3.56)	(5.84)	(1.44)
Year = 2001/02	-1.00**	-1.25**	-1.00	3.00	13.25*	2.84
	(0.28)	(0.40)	(1.13)	(3.56)	(5.84)	(1.44)
Year = 2002/03	-1.00**	-1.00*	1.75	15.25**	37.25**	7.16**
	(0.28)	(0.40)	(1.13)	(3.56)	(5.84)	(1.44)
Year = 2003/04	-0.25	0.25	5.50**	25.50**	37.50**	9.01**
	(0.28)	(0.40)	(1.13)	(3.56)	(5.84)	(1.44)
1998/99 in England	0.50	1.50*	2.50	-3.25	-16.75*	-0.63
	(0.40)	(0.56)	(1.59)	(5.04)	(8.26)	(2.04)
1999/00 in England	0.25	0.50	1.25	-9.25	-34.75**	-6.07**
	(0.40)	(0.56)	(1.59)	(5.04)	(8.26)	(2.04)
2000/01 in England	0.25	0.50	0.25	-11.75*	-42.75**	-9.11**
	(0.40)	**(0.56)**	**(1.59)**	**(5.04)**	**(8.26)**	**(2.04)**
2001/02 in England	**1.75****	**3.00****	**5.00****	**-1.50**	**-33.50****	**-5.83****
	(0.40)	**(0.56)**	**(1.59)**	**(5.04)**	**(8.26)**	**(2.04)**
2002/03 in England	**1.50****	**2.50****	**3.75***	**-11.25***	**-68.75****	**-13.01****
	(0.40)	**(0.56)**	**(1.59)**	**(5.04)**	**(8.26)**	**(2.04)**
2003/04 in England	**1.25****	**1.50***	**0.00**	**-27.75****	**-110.25****	**-23.96****
	(0.40)	**(0.56)**	**(1.59)**	**(5.04)**	**(8.26)**	**(2.04)**
Number of obs	56	56	56	56	56	56

Appendix B. Panel Data Estimation Procedure

The model as specified in Section 2.2 is given by

$$o_{ht} = \alpha + \beta d_{h,t-1} + z'_{ht}\theta + \sum_i \gamma_i o_{h,t-i} + \eta_h + v_{ht}$$

and direct estimation by Ordinary Least Squares would lead to biased estimates, due to the correlation of the lagged dependent variables $o_{h,t-j}$ and the hospital effect η_h. The fixed effects, or within groups, estimator is also biased, but this bias decreases with an increasing number of time periods.

The Arellano-Bond (1991) procedure is to transform the model into first-differences to get rid of the hospital fixed effects:

$$\Delta o_{ht} = \beta \Delta d_{h,t-1} + \Delta z'_{ht}\theta + \sum_i \gamma_i \Delta o_{h,t-i} + \Delta v_{ht} \ ,$$

where $\Delta o_{ht} = o_{ht} - o_{h,t-1}$. As now $\Delta o_{h,t-1}$ is clearly correlated with Δv_{ht}, one can use lags $o_{h,t-2},...,o_{h,t-2-j}$ as instruments employing an instrumental variables estimation technique, provided the idiosyncratic errors v_{ht} are not serially correlated. In the model for admissions, we find by means of the Hansen test for overidentifying restrictions that there is correlation between $\Delta v_{ht} = v_{ht} - v_{h,t-1}$ and $d_{h,t-2}$, therefore requiring instruments dated $d_{h,t-3},...,d_{h,t-3-j}$. No such problem was found for planned admissions, removals and suspensions, and therefore $d_{h,t-2},...,d_{h,t-2-j}$ can be used as instruments in those cases, improving the strength of the instruments, but still allowing for correlation between $v_{h,t-1}$ and $d_{h,t-1}$. To allow for measurement error in the workforce variable, this variable is treated as endogenous and instruments dated $z_{h,t-2},...,z_{h,t-2-j}$ are used.