

Accountability and incentives: The impacts of different regimes on hospital waiting times in England and Wales*

Timothy Besley, Gwyn Bevan and Konrad Burchardi
London School of Economics

September 25, 2008

Abstract

Improving accountability in public services has been a central objective of many public sector reforms in recent years. Chief among these have been efforts to generate observable performance measures as a basis for monitoring performance. This paper examines a natural experiment in regimes applied to waiting list targets for hospital admissions in England and Wales. Prior to 2001, each country had similar policies, organisational structures for hospital care, and levels of resources. After 2001, the principal difference between the countries were the consequences for hospitals that failed to meet targets for waiting times: in England, failure resulted in sanctions in a process of ‘naming and shaming’, but in Wales, failure was perceived to result in extra resources. We use hospitals in Wales as a ‘control group’, to examine the effect of ‘naming and shaming’ in England. We found that this policy did indeed reduce waiting times in England as compared with Wales. However, there is some evidence there was in England, initially, some shuffling of prospective patients to meet specific targets which increased mean waiting times.

*We are grateful to Oliver Bevan for constructing the database we have used and for comments on an earlier draft, and to staff in the Welsh Assembly Government for help in the supply of data from Welsh hospitals.

1 Introduction

Improving accountability in public services has been a central objective of many public sector reforms in recent years. Chief among these have been efforts to generate observable performance measures as a basis for monitoring performance. However, such efforts are not without controversy. Measurable performance criteria do not always reflect things which matter to consumers. Worse still, this can result in effort being directed away from desirable goals towards meeting the target as suggested in the multi-tasking model of Holmstrom and Milgrom (1991).

This paper contributes to an emerging body of literature that examines the consequences of efforts to enhance accountability by increasing performance measurement and using this to punish/reward providers (Heckman et al., 1997; Heinrich, 2002; Rosenthal and Frank, 2006; Burgess and Ratto, 2003; Doran et al., 2008; Campbell et al., 2007). The context is the introduction of a regime of ‘naming and shaming’ for failure to achieve waiting list targets for hospitals in the National Health Service (NHS) in England. By the year 2000, responsibility for running the NHS in England, Scotland, and Wales was devolved to governments of each country (devolution was largely stalled in Northern Ireland). Only the government in England sought to change the system of perverse incentives that had developed across the different countries: from one that ignored success and rewarded failure to one that celebrated success and penalised failure. This was done through the radical and controversial system of annual star rating of NHS organisations, between 2001 and 2005, which ‘named and shamed’ those that ‘failed’, which were zero rated; and offered ‘earned autonomy’ to the ‘high-performing’ three-star organisations. In Wales and Scotland the system of perverse incentives continued alongside the very different regime at work in England. The policy differences that have emerged following devolution offer a natural experiment to evaluate their impacts.

We estimate a difference-in-differences model of the proportion of people on the waiting list for different times in Wales and England at the level of a hospital trust.¹ Trust fixed effects allow us to control for sources of

¹Earlier papers have highlighted differences at the national level in performance on waiting times (Alvarez-Rosete et al., 2005; Bevan, 2006; Bevan and Hood, 2006a,b). Hauck and Street (2007) undertook a detailed analysis across three English hospital trusts and one Welsh hospital trust close to the English-Welsh border. Propper et al. (2008) estimated difference-in-differences models of the proportion of people on the waiting list who waited

unobserved heterogeneity and we also control for common shocks through the inclusion of year dummy variables. We exploit the fact that the timing and nature of the treatment in England and Wales is different to identify the effect of the target on waiting lists. The results show that targets were indeed effective in bringing down the waiting times in England, where, for a NHS trust, with a median number of patients waiting in June 1999, the estimated effect of the 18-, 15- and 12-month targets is to have reduced the numbers of patients waiting longer than the targeted time to zero. The 9-month target is estimated to have reduced the number of patients waiting between 9 and 12 months by 67%.

The remainder of the paper is organized as follows. The next section gives the background policy context for the analysis. It outlines the common development of the organisation and governance of the NHS in the UK, the different regimes that then developed in England and Wales after devolution, and what is known about their impacts. Section three outlines the data and methodology used while section four presents the results. The concluding comments are in section five.

2 Background and Context

The NHS in the UK was created in 1948 to provide universal coverage financed by taxation, largely free at the point of delivery in a publicly-organised system of functional units (acute hospitals) and units defined territorially (for example, care for the mentally ill, ambulances, primary care, dentistry), which broadly allowed clinical autonomy to medical professionals in their decisions on treating patients Klein (2006). Periodic reorganizations changed the boundaries, names and nature of those sub-units, but not the system's other abiding characteristics. Reorganization in the 1970s created health authorities in England and Wales in hierarchical structures that were responsible for planning services for defined resident populations and running hospitals and community health services.²

From its inception, the prevailing view was that the NHS was staffed by publicly spirited workers who needed no incentives, sanctions or rewards –

over 6, 9 and 12 months in England and Scotland. All these analyses strongly suggest that the policy of star ratings did reduce waiting times in England.

²Different legislation applied the same principles with the creation of Health Boards in Scotland and Health and Social Service Boards in Northern Ireland.

see Le Grand (2003) for further discussion. But this view began to change over time. For example, Enthoven (1985) claimed serious problems with the hierarchical organization of the NHS and its lack of incentives. He described the NHS as being in a ‘gridlock’: ‘caught in the grip of forces that make change exceedingly difficult to bring about’ (p. 9), the fundamental problem being that ‘the system contains no serious incentives to guide the NHS in the direction of better quality of care at reduced cost’ (p .13). He recommended the introduction of incentives by requiring providers to compete in an ‘internal market’.

This view was ultimately influential in shaping the Thatcher government’s pursuit of reform. In response, the so-called ‘internal market’, based on the principle of provider competition, was implemented between 1991 and 1997, with a funding system that promised that ‘money would follow the patient’(Secretaries of State for Health, Wales, Northern Ireland and Scotland, 1989; Bevan and Robinson, 2005; Klein, 2006). This led to the reorganization of the health authorities into purchasers, which contracted for hospitals and community health services.

In spite of these reforms, waiting times remained a problem.³ In response, the government in England announced new policies in 2000, with an objective of cutting maximum waiting times for elective admission from 18 months to 6 months by the end of 2005 (Secretary of State for Health, 2000). The principal policy instrument for delivering this transformation was the system of ‘star ratings’, which applied to acute hospitals from 2001 to 2005 (Department of Health, 2001, 2002; Commission for Health Improvement, 2003a,b; Healthcare Commission, 2004, 2005). This process gave each organization a score from zero to three stars based on performance against a small number of ‘key targets’ and a larger set of targets and indicators in a ‘balanced scorecard’. Organizations that failed against ‘key targets’, and were ‘zero-rated’, were ‘named and shamed’ as ‘failing’, and their chief executives were at risk losing their jobs: this happened to six chief executives of the 12 trusts given ‘zero rating’ in 2001 and four of these improved their rankings in

³Failing providers do not exit the market (Tuohy, 1999; Enthoven, 1999; Secretary of State for Health, 2000; Bevan and Robinson, 2005). It has also proven difficult to create an effective demand-side either by commissioning services through purchasing organisations or patient choice. The evidence from two systematic reviews (Marshall et al., 2003; Fung et al., 2008) of the literature on the effects of publishing information on hospital performance found that patients did not respond as consumers to use evidence on hospital performance to switch from poor to good hospitals.

the following year's star ratings (Beverley and Haynes, 2005). Organizations that performed well on both the 'key targets' and the 'balanced scorecard', and achieved the highest rating of three stars, were rewarded by being publicly celebrated for being 'high performing' and granted 'earned autonomy' (Bevan and Hood, 2006a,b).

In the models used for star ratings, the 'key targets' were most important. To justify the claim that star ratings offered a rounded assessment of performance, key targets were supplemented by a wider set (about forty) targets and indicators in a so-called 'balanced scorecard'. Within the star ratings of acute trusts and PCTs, reducing hospital waiting times was of overriding importance; failure to deliver these targets could result in being zero-rated. For acute trusts: six of the nine key targets were for waiting times (the other three were achieving a financial balance, hospital cleanliness, and improving the working lives of staff): and one of the three domains in 'balanced scorecard' was the 'patient focus', which was also dominated by waiting time targets. The star rating for Primary Care Trusts also included three key targets for waiting times. Table 1 gives the targets for waiting for elective admission in England showing how these became more demanding over the five years of star rating.

The application of targets became more explicit as the system developed. In the first year (2000/01) the 18-month target applied at end of March only. In the second year (2001/02) the targets were set were that 'no patients waiting more than 18 months for inpatient treatment' and 'fewer patients waiting more than 15 months for inpatient treatment'. From the third year, failure was defined in terms of the number of breaches and these for each year were as follows

- 2002/03: the sum of the number of patients waiting longer than 15 months at the end of each the first 11 months of 2002/03 plus the number of patients who were waiting longer than 12 months at the end of March 2003;
- 2003/04: the sum of the number of patients waiting longer than 12 months at the end of each the first 11 months of 2003/04 plus the number of patients who were waiting longer than 9 months at the end of March 2004;
- 2004/05: the sum of patients waiting more than 9 months at each month from April 2004 to March 2005.

The ‘star rating’ system succeeded in conveying to those who worked in the NHS that reducing waiting times mattered by ‘naming and shaming’ those that failed. The evidence from the US is that systems of performance assessment that are designed to inflict reputational damage on poorly performing hospitals have an impact where markets do not (Hibbard et al., 2003, 2005a; Chassin, 2002; Bevan and Hamblin, 2009). Hibbard identified the four requisite characteristics for a system to inflict damage: these are that it be a ranking system, published and widely disseminated, easily understood by the public, and followed up by future reports. The ‘star rating’ system satisfied all these characteristics (Mannion et al., 2005).

In contrast with England, following devolution, the government in Wales initially abandoned targets for waiting times (Hauck and Street, 2007), and when these were introduced from 2001, a report from the Auditor General for Wales (2005, p. 36) observed that, although waiting times were ‘an important part of the Welsh Assembly Government’s overall health policy. Waiting time targets have been set out in a variety of documents and not always been clearly and consistently articulated or subject to clear and specific timescales’. We rely on that report for understanding of the changing policy in Wales on reducing waiting times.

Targets for waiting times were used in Wales more as an aspiration, in the hope that managers would respond. These targets were adjusted to reflect variations in local circumstances, with some Trusts allowed a number of breaches, which were not publicized, so people on these waiting lists would have been misled to expect treatment within the relevant waiting time target (Auditor General for Wales, 2005, p. 35). The system of reporting performance in Wales from 2003/04 was through targets specified through the Service and Financial Framework (SaFF) but there was confusion over the relative priority of the various SaFF targets (although Trusts perceived financial and waiting time targets to be more important than others); which was exacerbated by the large number of targets Trusts were expected to achieve (104 in 2003-04, although these were reduced to 40 in the following year) (Auditor General for Wales, 2005, p. 39). There was a website that indicated to the public likely waiting times by specialty, hospital, and specialist (Health of Wales Information Service, 2006), but there was no equivalent system to star ratings in Wales. There was no ranking system, no attempt to inform the public about hospitals’ performance against targets through regular reports. Whereas in England, the governments’ response to the problem of long waiting times was to set ambitious targets, in Wales, targets were

set to reflect existing poor performance. This is illustrated in Table 2, which gives the targets in place in 2005, the final year of star ratings in England. The Auditor General for Wales (2005, p. 17) also commented on the contrast between the ambitious target set in England for a pathway-based maximum waiting time of 18 weeks from GP referral to treatment, to be achieved by 2008, whereas the Welsh Assembly Government had ‘no similarly clear strategy outlining how it intends to reduce target waiting times over the medium term’.⁴

In addition to the reforms described here that affected the operation of the NHS, there were also increases in funding with NHS expenditure being increased by 5 per cent in real terms over the six years from 2001-02 (Smee, 2005). However, it is important to observe for the exercise undertaken here that funding levels for both England and Wales were similar over this period. Thus, since so much else was similar in the NHS in England and Wales, the principal difference is in the governance regime.

There is now a large theoretical literature looking at why organizations that cohere around a public service motive may be different from standard private organizations run to maximize profit. Here is not the place to review that literature in detail. However, it is useful to outline how some of the ideas in that literature affect the interpretation of the results developed here.

A key difficulty in achieving accountability and improving incentives in public services is the difficulty of measuring the ‘quality’ of the output in a relevant sense. Public services generally run on the basis of some kind of non-profit mission as discussed in Besley and Ghatak (2003) where mission is defined by Wilson (1989, p. 95), as a culture ‘that is widely shared and warmly endorsed by operators and managers alike’. This measurement problem leads government to develop broad measurable indicators which are then used to regulate the performance of public service providers. Some aspects of accountability can then be tied directly to such measurable indicators.

⁴A complication in comparing performance in England and Wales is that from 1 April 2004, the government in Wales introduced the ‘second offer scheme’ for patients on the inpatient and day case waiting list if they had waited, or were likely to wait, over 18 months, or would breach the specific targets for particular treatments. This scheme paid for such patients to be treated at alternative providers (private hospitals in Wales or hospitals in England) at no charge to the hospital for these patients as the costs were paid from made central funds. This scheme was extended in June 2004, so that, by March 2005, it would guarantee an offer of treatment by an alternative provider for those waiting over twelve months (Auditor General for Wales, 2005, p. 9).

Since Baker (1992) and Holmstrom and Milgrom (1991), it has been appreciated in the theoretical literature that care needs to be taken in using imperfect performance indicators to regulate the operation of organizations. Using high-powered incentives for observable performance can be problematic in this context. Even if you get more of what you are rewarding, as you would expect, it is essential that this does not come at the expense of poorer performance on other, harder-to-measure, dimensions. This effort diversion is frequently referred to as ‘gaming’ in the literature on public sector performance : Smith (2005) offers a typology; its problematic existence has been recognised in empirical studies with financial incentives (see for example, Heinrich, 2002; Doran et al., 2008; Burgess and Ratto, 2003); and Bevan and Hood (2006a,b) have shown how ‘naming and shaming’ also resulted in gaming.

3 Data and Methodology

This section discusses the data and the way in which we use these to construct a test for the impact of waiting time targets on the length of waiting times.

3.1 Data

We obtained data on the distribution of waiting times for each NHS trust in Wales and England.⁵ The data is a snapshot of the hospitals’ waiting lists on the last day of each financial quarter of the NHS. The length of the waiting time is classified in 7 different 3-month bands (‘waiting between 0 and 3’, ‘between 3 and 6 months’ etc. with the highest being ‘waiting more than 18 months’) and our data consists of the number of patients waiting in each of those bands.⁶ It covers 28 quarters in the period from the first quarter of the financial year 1999/2000, corresponding to end of June 1999, to the last quarter of the financial year 2005/2006, corresponding to end of March 2006.

The waiting list statistics are patients waiting to be admitted either as a day case or ordinary admission. The principal difference in definitions be-

⁵The data can be downloaded at www.performance.doh.gov.uk/waitingtimes/index.htm and www.statswales.wales.gov.uk/ReportFolders/ReportFolders.aspx. We accounted for mergers by summing the data for the merged hospitals prior to the merger.

⁶For example, in any hospital we have data on the number of patients waiting between 9 and 12 months on the last day of any financial quarter.

tween Wales and England is that, in Wales, all referrals are included whatever the source, whereas in England, only referrals from medical and dental general practitioners are included (Auditor General for Wales, 2005, pp. 50-53).

To get a feel for what the data show, Figures 1-7 present the sum of patients waiting per region for each of the waiting bands. Table 3 presents the mean and median of the number of patients waiting per trust in each waiting band for the 9 regions in our data, i.e. Wales and the 8 English regions. It is evident from these figures that waiting lists fell in line with the targets and that the gap with Wales opened over the period, suggesting that the targets did have an impact on hospital policy in England.

3.2 Methodology

To evaluate the effect of the English regime of ‘naming and shaming’ for failure to achieve targets for waiting times for hospital admission we make use of the fact that around the time when this regime was introduced in England, although targets were introduced into the NHS in Wales, this was without a regime of ‘naming and shaming’. In section 4 we will come back to this when we assess the robustness of our main results. Given that the Welsh and English NHS are otherwise organisationally similar and subjected to the same funding, we believe the Welsh NHS to be a suitable control group for evaluating the ‘treatment’ of the English NHS. The effect of each target can then be identified by running for each waiting band $w = 0, 3, 6, 9, 12, 15, 18$ a simple difference-in-difference specification of the form

$$y_{itw} = \beta_w \cdot \text{target}_{itw} + \delta_{tw} \cdot D_t + \gamma_{iw} \cdot D_i + \epsilon_{itw} \quad (1)$$

where D_t is a matrix of time dummies, D_i are NHS trust dummies and target_{itw} a dummy being 1 if hospital i is in a region where at time t a target for waiting category w existed.

4 Results

4.1 Core Results

Focusing first on the effect of the waiting time targets on the targeted waiting category, we present results from the specification in equation (1). The raw data show that there were negligible numbers of patients in the English NHS

waiting longer than each target, for 18, 15, 12 and 9 month, when it came into force. Table 4 presents the coefficient estimates of the effect of each of these targets in the English NHS on the number of patients waiting over 18 months, between 15 and 18 months, between 12 and 15 months and between 9 and 12 months, respectively. Our results show that in the first year of this regime, NHS Trusts sought to achieve the 18-month target only. Subsequently, they sought to achieve the target for that year and make progress towards future targets. This is what we would expect, for a new regime with increasingly demanding targets. The 18-month target had been in place for the NHS in England since 1995 (NHS Executive, 1995). What was new about the regime was that sanctions applied for failure to meet that target in 2001. Experience of hitting (or missing) that target would have made it clear that systemic changes would be necessary to continue to meet future targets.

In an English hospital with a median number of patients waiting in June 1999, the first three targets are estimated to have reduced the numbers of patients waiting longer than the targeted time to zero. The 9-month target is estimated to have reduced the number of patients waiting between 9 and 12 months in an English NHS trust by 67%, again compared to the median number of patients waiting in June 1999. The results in table 5 also suggest such an early treatment effect in anticipation of the announced target. Table 5 presents similar regressions to table 4, but previous treatments are now included. Column (4) suggests that the number of patients waiting between 9 and 12 months in the English NHS decreased already significantly at the times when the 12 and 15-month targets were enacted, so in the two years before the 9-month target actually came into force. Including this early treatment effect, the 9-month target's estimated effect is to have achieved that no patients were waiting more than 9 months in a median English NHS trust. The early treatment effect that we described above is particularly clear for the 12-month target. Column (3) also shows that the number waiting longer than 12 months already dropped in the two years prior to the 12-month target coming into force. However, for the earlier 15-month target no significant prior drop in the waiting list is estimated.

Taken together, these results suggest that the targets were effective in reducing long waits. They suggest as well that hospitals early on managed their waiting lists to fulfill the later targets. However, they do not suggest that the hospitals achieved this by treating more patients. In fact the targets' effects might have resulted from a different management of the waiting lists. Indeed, the results of table 5 show that the reduction in long waits came

at the expense of an increase in the numbers of patients waiting for shorter time periods. In particular, columns (5)-(7) show that the number of patients waiting between 3 and 9 months significantly increased both after the introduction of the 15- and the 18-month target. For the interpretation of these numbers it is important to recall that we use census rather than discharge data. If the hospitals had reacted to the targeting regime by treating additional patients once they waited for 9 months, and who would have waited even longer prior to the targeting regime, the numbers of patient waiting less than 9 months should not change. The increase in the waits between 3 and 9 months hence shows that patients who, in the absence of the targeting regime, would have waited 0 to 6 months, were now left waiting until their waiting time approached the maximum allowed.

Again the later targets, i.e. the 12- and the 9-month target, did not have this effect. Their coefficient estimates are not significantly positive in columns (5)-(7). But they are generally non-negative and never significantly so. This indicates that while the number of close-to-9-month waits did not increase further, the hospitals were not able (or had no incentive) to cut back the previously increased level of close-to-9-month waits either. Further, the estimated overall effect of the targeting regime, measured by the sum of the effects of the four targets, is to have increased the number of patients waiting in all three categories below 9 months waiting time. A t-test for this sum is significant least at the 5% level for all three categories.

The exception to this rule is the effect of the 9-month target on the number of patients waiting between 6 and 9 months. However, considering the early treatment effects outlined above, this might well be driven by the later to be introduced 6-month target.^{7 8}

Taken together the results of table 5 seem to suggest that the targets were effective in reducing long waits, but this was done, at least in part, not by treating more patients, but by prioritising the treatment of patients waiting for a long time, which and consequently increased the mean waiting time.

⁷This was to be achieved by December 2005 under the new regime of ‘annual healthchecks’, the successor to star ratings.

⁸To check the robustness of those results we allowed for regional time trends and considered the Welsh targets as treatment after the introduction of the second offer scheme in March 2004 (see section 4.2). All of our results remain the same.

4.2 Robustness

As outlined above, although Wales later introduced targets for reducing long waits, these targets were not strictly defined and there were no sanctions for failure to achieve them. Hence in the previous section we did not consider the effects of targets in Wales. In order to check the robustness of our results to our interpretation of the Welsh targeting regime table 6 presents equivalent estimation results as table 5, with the difference that we now define the Welsh second-offer scheme to patients waiting more than 18 and later 12 months as treatment from April 2004 and April 2005, respectively, onwards. It can be seen that our results do not change qualitatively. The estimated effects of the targets are still negative, large and significant. The waiting numbers in the 9-12 and 12-15 month categories drop well before the target and we do not observe a significant early treatment effect for the 15-month target. The targets' effects on the number of patients waiting between 0 and 9 months are generally non-negative and often significant. The only difference to the previously presented results is the effect of the 9-month target. Its effect at the time of the implementation is not significant. The effect of the target seems to have been a reduction of the patients waiting 9-12 months well before the target came into force. Secondly, the 9-month target is estimated to have significantly increased the number of patients waiting between 0 and 6 months and not have reduced the number of patients waiting between 6 and 9 months. This adds to the evidence that part of the reduction in long wait was achieved by a different management of the waiting lists. The overall effect of the targeting regime on the number of patients waiting between 0-3, 3-6 and 6-9 months is again estimated to be positive and significant.

4.3 Evidence of Gaming?

Kelman and Friedman (2007) examined various potential types of gaming in response to another 'key target' for waiting times in England: this concerned patients being seen and treated within four hours in Accident and Emergency Departments (A&E Departments - known as emergency rooms in the US). There is evidence of dramatic improvements in England in meeting this target and also of gaming (Bevan and Hood, 2006a,b). Kelman and Friedman examined whether gaming occurred in shuffling patients in order to meet the four-hour target with consequences of a decrease in the percentage of patients treated within two hours and an increase in mean waiting time. They

found evidence that the opposite occurred: shorter waits were associated with lower mean wait times and a higher fraction of patients treated in under two hours. This seems to us a poor test of gaming, as the A&E four-hour wait target was about a fundamental change in the culture and organisation of these departments from the tradition of triage (which we have been told was introduced around the time of the first world war). That is to say that A&E departments worked on the principle that life-threatening emergencies required urgent treatment and it did not matter how long the others waited. To implement the four-hour target it is simplest to introduce a good system for all patients than shuffling people around according to their waiting time. A much better test of this type of gaming is whether mean waiting time increased in England in relation to the target for elective admission for 2001, that no patient should wait for more than eighteen months after having been referred by a general practitioner, which does look vulnerable to shuffling of patients.

We now test for how far hospitals may have tried to game targets by shuffling patients across different categories of waiting times. We do this, by first seeing the overall impact on mean waiting times and then produce a counterfactual of what would have happened had we not seen any increase in waiting lists at other time lengths. To benchmark this possibility, table 7 presents different specifications of how the mean-waiting time changed with the introduction of the four targets under study. The table shows that mean waiting times did indeed decrease after the introduction of the 15- and 12-month targets. However, they may have increased after the introduction of the 9-month target.

We now compare this with a hypothetical mean waiting time constructed from Table 5. Specifically, we use the coefficients in Table 5 after having set the coefficients in columns (5) through (7) to zero. This assumes that targets had no effect on the distribution of below-9-month waits. The result of this comparison is summarized in Figure 8, which plots the actual mean waiting time in England alongside the hypothetical mean waiting time.⁹ This calculation suggests in the first two years of the star rating regime (quarters 8 to 16), the mean-waiting time would have been up to one month shorter had there not been any change in waiting at other time lengths; this fell to

⁹Since the hypothetical mean waiting time is calculated from the predicted values of the regressions in table 5, we present as well the mean waiting time calculated from the predicted values. It follows the actual mean waiting time closely.

six months in the third year (quarters 16 to 20), and subsequently to zero in the fourth year. This does suggest that there was initially some gaming of the targets, which had a material impact on average waiting times, but this declined and ceased altogether by the fourth year.

While this exercise is constructive, it does not get at wider possibilities for redeploying resources to meet the targets which had detrimental effects on patient care – this would be the classic multi-tasking behavioural response. We can offer no evidence for or against the hypothesis that other dimensions of patient care were affected by waiting targets. But there is little evidence that the longer waiting times in Wales were offset by improvements in other areas. Hauck and Street (2007) report results of a detailed analysis of four hospitals (three in England and one in Wales) which were close to the border and serving similar populations over the period from 1997/98 to 2002/03. In the English hospitals there was increased activity and low or declining mortality rates; but the Welsh hospital had no increased activity and high and rising mortality rates. Leatherman and Sutherland (2003), report mortality rates to have been higher in Wales than in England from: causes considered amenable to healthcare, coronary heart disease, stroke and diabetes. The Royal College of Physicians (2006) found that patients in Wales were more likely die from stroke, or if they survived would have higher levels of disability than in England or Northern Ireland (Royal College of Physicians, 2006). Further examination of the question whether reducing waiting times in England had other adverse effects in comparison with Wales is an important avenue for future investigation.

5 Conclusion

This paper has exploited a natural experiment between two regimes for hospital waiting time targets: ‘naming and shaming’ in England for failure and rewarding failure in Wales. Using Wales a control group, we found that ‘naming and shaming’ did reduce the time that patients waited. In fact, such waiting has all but been eliminated by the use of targets combined with real sanctions for hospital chief executives.

Given that the identification proposed here is quite clean, it is reasonable to argue that what we have found is a behavioural effect at the hospital level. It shows that targets with sanctions – part of the naming and shaming regime that has been used in recent years to improve public services in England – has

had an impact. We are still left with questions over the extent to which the dramatic reductions in waiting times in England were funded by the massive and sustained increases in funding, or by provider rents, or by a worsening of performance in areas that were not targeted. The available evidence suggests that increased funding together with ‘naming and shaming’ meant that the performance of the NHS in England (as measured by waiting times) was transformed; and the absence of ‘naming and shaming’ meant that no similar transformation took place in Wales. And that providers in Wales were able to use the extra funding to extract provider rents, particularly as we were unable to find any concrete evidence of changes in England in response to the regime there were detrimental to patient welfare as compared to Wales. In the next phase of empirical research in this area, it will be interesting to look for other consequences of the regime of ‘naming and shaming’.

References

- Alvarez-Rosete, Arturo, Gwyn Bevan, Nicholas Mays, and Jennifer Dixon**, “Effect of diverging policy across the NHS,” *BMJ*, 2005, *331* (7522), 946–950.
- Auditor General for Wales**, *NHS waiting times in Wales. Volume 1 - The Scale of the problem*, Cardiff: The Stationery Office, 2005.
- Baker, George P.**, “Incentive Contracts and Performance Measurement,” *Journal of Political Economy*, June 1992, *100* (3), 598–614.
- Besley, Timothy and Maitreesh Ghatak**, “Incentives, Choice, and Accountability in the Provision of Public Services,” *Oxford Review of Economic Policy*, 2003, *19* (2), 235–249.
- Bevan, Gwyn**, “Setting Targets for Health Care Performance: Lessons from a Case Study of the English NHS,” *National Institute Economic Review*, 2006, *197* (1), 67–a–79.
- **and Christopher Hood**, “Have targets improved performance in the English NHS?,” *British Medical Journal*, 2006, *332* (7538), 419–422.
- **and –**, “What’s measured is what matters: targets and gaming in the English public health care system,” *Public Administration*, August 2006, *84* (3), 517–538.
- **and R. Hamblin**, “Hitting and missing targets by ambulance services for emergency calls: impacts of different systems of performance measurement within the UK,” *Journal of the Royal Statistical Society*, 2009, *172* (1), 1–30.
- **and Ray Robinson**, “The Interplay between Economic and Political Logics: Path Dependency in Health Care in England,” *Journal of Health Politics Policy and Law*, 2005, *30* (1-2), 53–78.
- Beverley, C. and J. Haynes**, *Franchised Trusts*, Health Management Specialist Library: Management Briefing, NeLH Health Management Specialist Library, 2005.

- Burgess, Simon and Marisa Ratto**, “The Role of Incentives in the Public Sector: Issues and Evidence,” *Oxford Review of Economic Policy*, 2003, 19, 285–300.
- Campbell, Stephen, David Reeves, Evangelos Kontopantelis, Elizabeth Middleton, Bonnie Sibbald, and Martin Roland**, “Quality of Primary Care in England with the Introduction of Pay for Performance,” *New England Journal of Medicine*, 2007, 357 (2), 181–190.
- Chassin, Mark R.**, “Achieving And Sustaining Improved Quality: Lessons From New York State And Cardiac Surgery,” *Health Affairs*, 2002, 21 (4), 40–51.
- Commission for Health Improvement**, *NHS Performance Ratings. Acute Organisations, Specialist Organisations, Ambulance Organisations 2002/03*, London: The Stationery Office, 2003.
- , *NHS Performance Ratings. Primary Care Organisations, Mental Health Organisations, Learning Disability Organisations 2002/03*, London: The Stationery Office, 2003.
- Department of Health**, *NHS performance ratings acute trusts 2000/01*, London: Department of Health, 2001.
- , *NHS performance ratings acute trusts, specialist trusts, ambulance trusts, mental health trusts 2001/02*, London: Department of Health, 2002.
- Doran, Tim, Catherine Fullwood, David Reeves, Hugh Gravelle, and Martin Roland**, “Exclusion of Patients from Pay-for-Performance Targets by English Physicians,” *New England Journal of Medicine*, 2008, 359 (3), 274–284.
- Enthoven, A.**, *In Pursuit of an Improving National Health Service*, London: Nuffield Trust, 1999.
- Enthoven, A. C.**, *Reflections on the Management of the NHS*, London: Nuffield Provincial Hospitals Trust, 1985.
- Fung, Constance H., Yee-Wei Lim, Soeren Mattke, Cheryl Damberg, and Paul G. Shekelle**, “Systematic Review: The Evidence That Publishing Patient Care Performance Data Improves Quality of Care,” *Annals of Internal Medicine*, 2008, 148 (2), 111–123.

- Grand, J. Le**, *Motivation, agency and public policy: Of knights and knaves, pawns and queens*, Oxford: Oxford University Press, 2003.
- Hauck, Katharina and Andrew Street**, “Do targets matter? A comparison of English and Welsh National Health priorities,” *Health Economics*, 2007, 16 (3), 275–290. available at <http://ideas.repec.org/a/wly/hlthec/v16y2007i3p275-290.html>.
- Health of Wales Information Service**, *Waiting Times Information* 2006.
- Healthcare Commission**, *2004 Performance Rating*, London: The Stationery Office, 2004.
- , *NHS performance ratings 2004/2005*, London: Healthcare Commission, 2005.
- Heckman, James, Carolyn Heinrich, and Jeffrey Smith**, “Assessing the Performance of Performance Standards in Public Bureaucracies,” *The American Economic Review*, 1997, 87 (2), 389–395.
- Heinrich, Carolyn J.**, “Outcomes-based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness,” *Public Administration Review*, 2002, 62 (6), 712–725.
- Hibbard, Judith H., Jean Stockard, and Martin Tusler**, “Does Publicizing Hospital Performance Stimulate Quality Improvement Efforts?,” *Health Affairs*, 2003, 22 (2), 84–94.
- , – , and – , “Hospital Performance Reports: Impact On Quality, Market Share, And Reputation,” *Health Affairs*, 2005, 24 (4), 1150–1160.
- , – , and – , “It Isn’t Just about Choice: The Potential of a Public Performance Report to Affect the Public Image of Hospitals,” *Medical Care Research and Review*, 2005, 62 (3), 358–371.
- Holmstrom, Bengt and Paul Milgrom**, “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design,” *Journal of Law, Economics, & Organization*, 1991, 7, 24–52.

- Kelman, S. and J. N. Friedman**, *Performance Improvement and Performance Dysfunction: An Empirical Examination of Impacts of the Emergency Room Waittime Target in the English National Health Service*, Cambridge: John F. Kennedy School of Government, 2007.
- Klein, R. E.**, *The New Politics of the National Health Service (5th ed)*, Oxford: Radcliffe Press, 2006.
- Leatherman, Sheila and Kim Sutherland**, *The Quest for Quality in the NHS: A mid-term evaluation of the ten-year quality agenda*, London: The Stationery Office, 2003.
- Mannion, Russel, Huw Davies, and Martin Marshall**, *Cultures for Performance in Health Care*, Maidenhead: McGraw Hill, 2005.
- Marshall, Martin N., Paul G. Shekelle, Huw T. O. Davies, and Peter C. Smith**, “Public Reporting On Quality In The United States And The United Kingdom,” *Health Affairs*, 2003, 22 (3), 134–148.
- NHS Executive**, “Revised and expanded Patients Charter: implementation,” *Health Service Guidelines HSG(95)13*, 1995.
- Propper, C., M. Sutton, C. Whitnall, and F. Windmeijer**, “Did ‘Targets and Terror’ Reduce Waiting Times in England for Hospital Care?,” *The B.E. Journal of Economic Analysis & Policy*, 2008, 8 (2 (Contributions)).
- Rosenthal, Meredith B. and Richard G. Frank**, “What Is the Empirical Basis for Paying for Quality in Health Care?,” *Medical Care Research Review*, 2006, 63 (2), 135–157.
- Royal College of Physicians**, *National Sentinel Stroke Audit*, London: Royal College of Physicians, 2006.
- Secretaries of State for Health, Wales, Northern Ireland and Scotland**, *Working for patients [CM 555]*, London: HMSO, 1989.
- Secretary of State for Health**, *The NHS plan [CM 4818-I]*, London: The Stationery Office, 2000.
- Smee, C.**, *Speaking Truth to Power: Two Decades of Analysis in the Department of Health*, Oxford: Radcliffe Press, 2005.

Smith, Peter C., “Performance Measurement in Health Care: History, Challenges and Prospects,” *Public Money & Management*, 08 2005, 25 (4), 213–220.

Tuohy, C., *Accidental Logics. The Dynamics of Change in the Health Care Arena in the United States, Britain and Canada*, New York: Oxford University Press, 1999.

Wilson, J. Q., *Bureaucracy: What Government Agencies Do and Why They Do It*, New York: Basic Books, 1989.

Figure 1

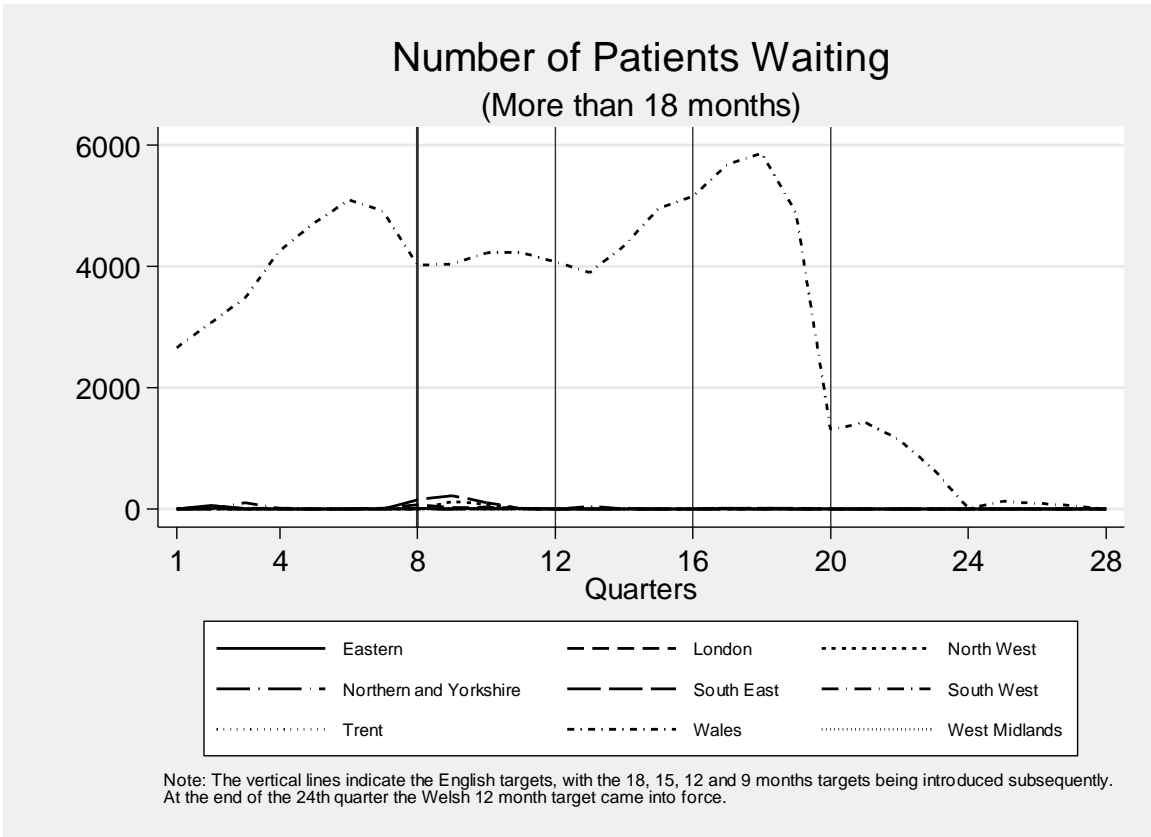


Figure 2

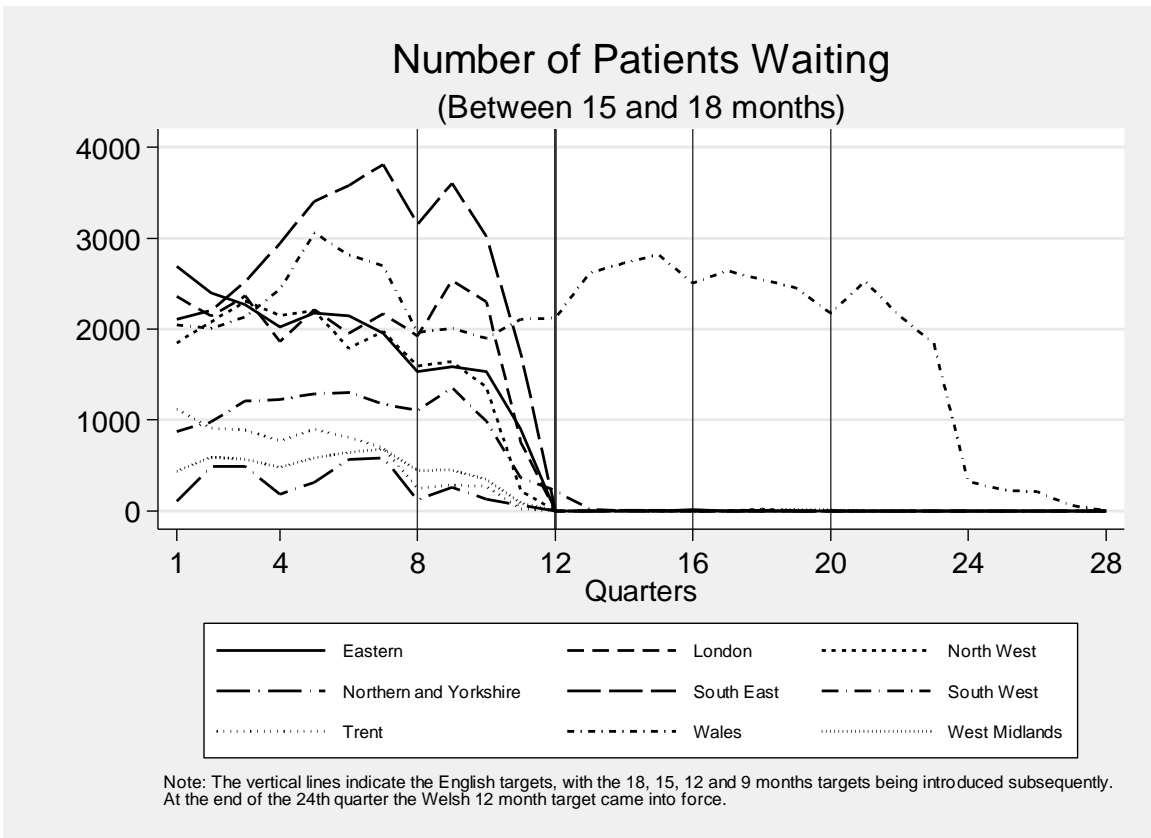


Figure 3

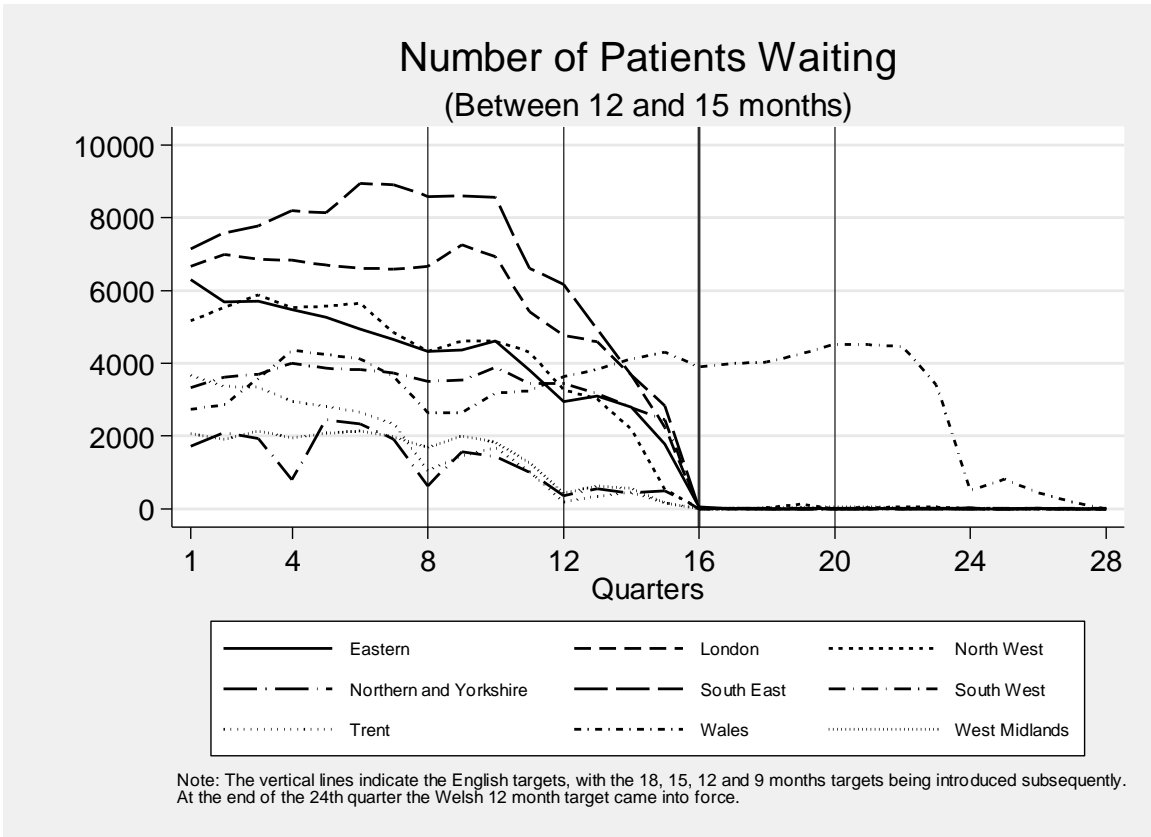


Figure 4

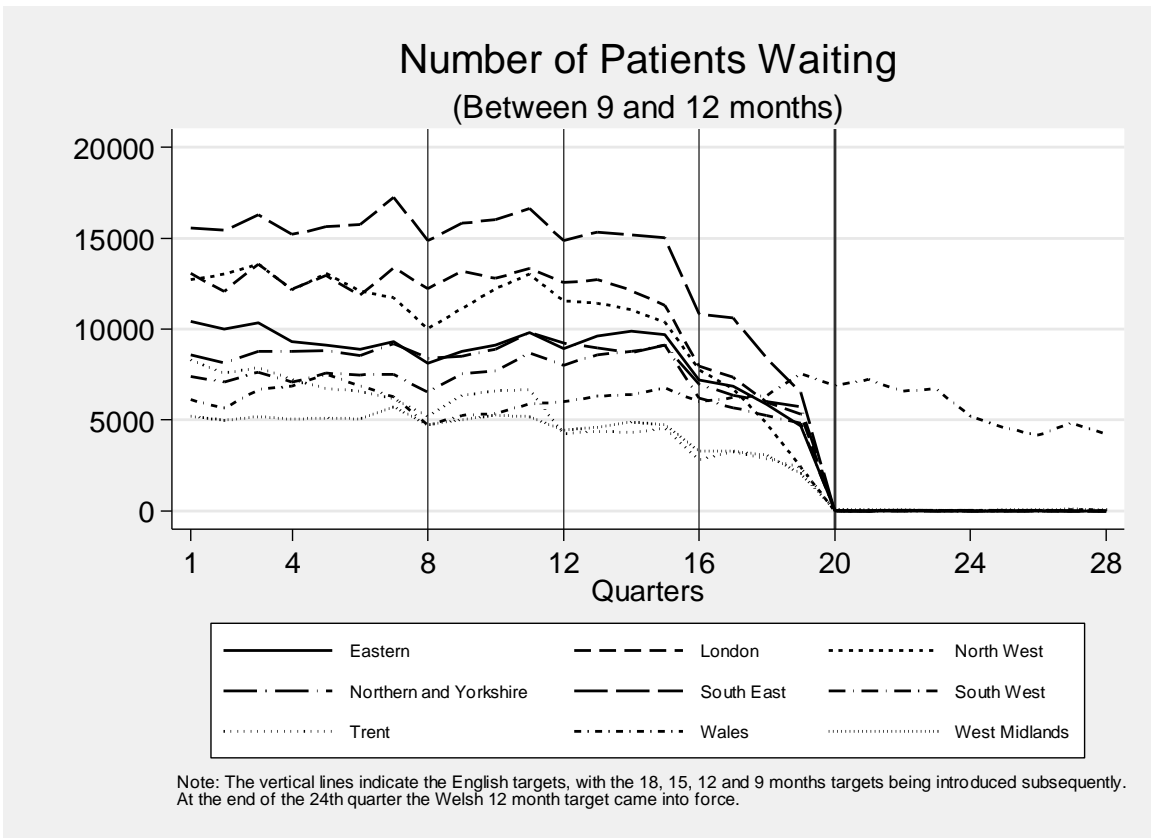


Figure 5

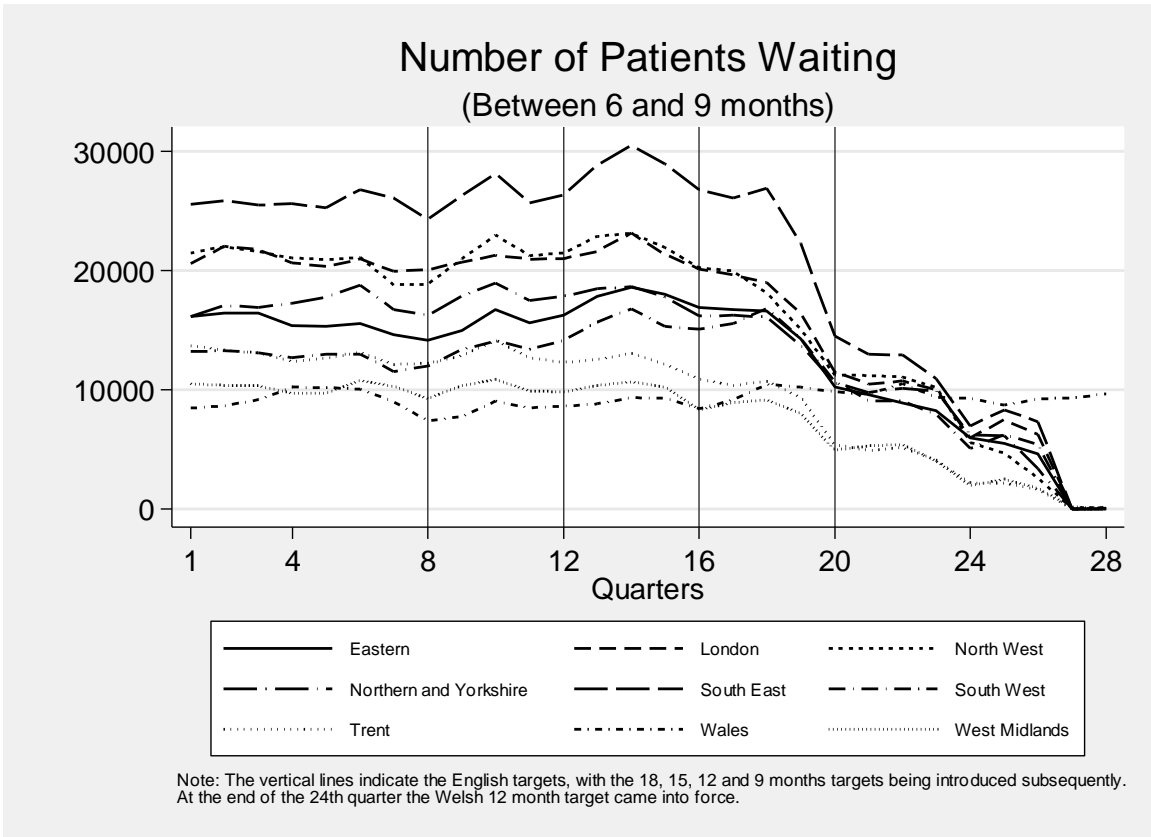


Figure 6

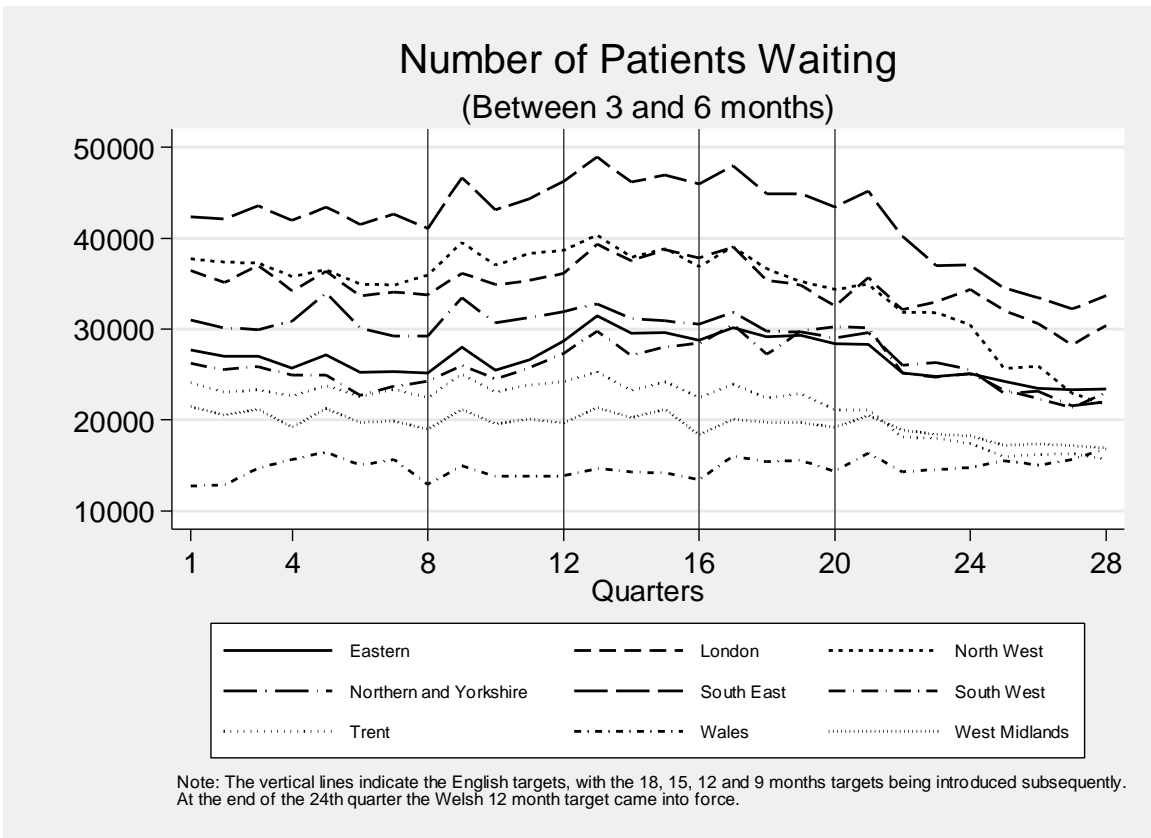


Figure 7

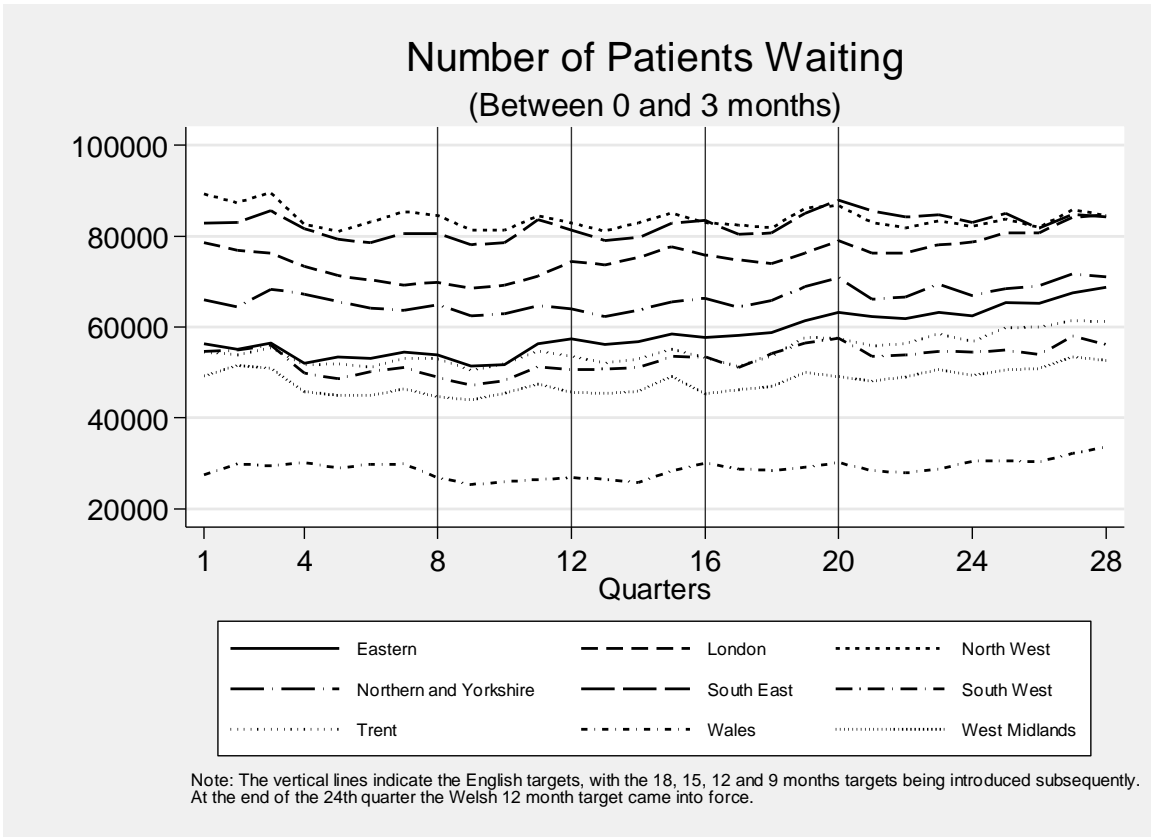


Figure 8

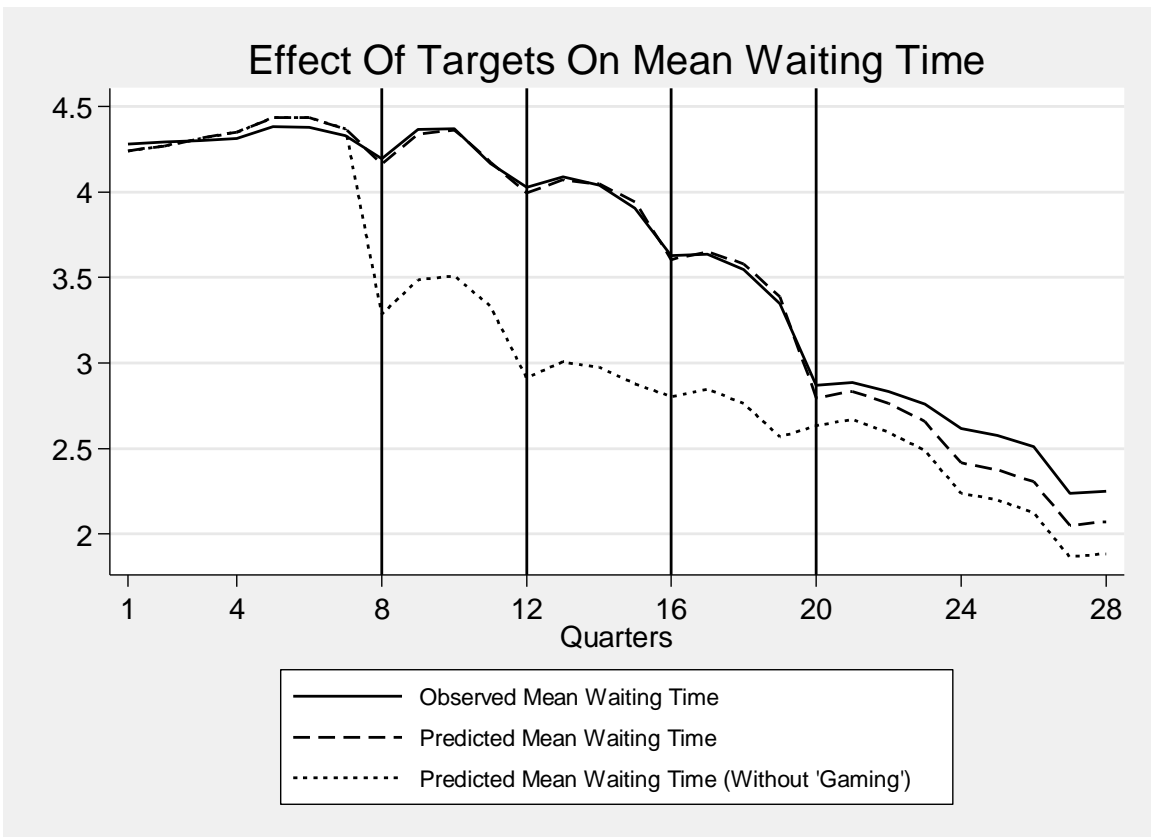


Table 1: Targets for waiting for elective admission in England

Year	Start of year (months)	End of year (months)
2000/01	18	18
2001/02	18	15
2002/03	15	12
2003/04	12	9
2004/05	9	9

Sources: Department of Health (2001, 2002); Commission for Health Improvement (2003a), Healthcare Commission (2004, 2005).

Table 2: Waiting time targets for England and Wales in 2005

Type of waiting	England (weeks)	Wales (weeks)
For first outpatient appointment	13	78
For inpatient / day case treatment	26	78

Source: Auditor General for Wales (2005a: 15)

Table 3: Mean and median numbers of patients waiting in the 9 regions

	No. of trusts	Mean (median) number of patients waiting per trust						
		0m<&<3m	3m<&<6m	6m<&<9m	9m<&<12m	12m<&<15m	15m<&<18m	18m<
Eastern	18	3128 (2856)	1540 (1493)	897 (796)	578 (468)	351 (252)	149 (74)	0 (0)
London	28	2804 (2962)	1300 (1291)	736 (648)	466 (369)	238 (151)	84 (67)	0 (0)
North West	22	4057 (3590)	1715 (1676)	976 (939)	578 (574)	235 (161)	84 (45)	0 (0)
Northern and Yorkshire	17	3877 (3372)	1822 (1699)	948 (1006)	504 (395)	101 (38)	6 (0)	0 (0)
South East	25	3313 (3112)	1695 (1563)	1022 (1049)	622 (591)	286 (240)	84 (54)	0 (0)
South West	18	3034 (2694)	1458 (1468)	733 (803)	410 (272)	185 (107)	49 (18)	0 (0)
Trent	14	3895 (3376)	1721 (1457)	978 (925)	593 (539)	262 (134)	80 (42)	0 (0)
West Midlands	17	2894 (2666)	1263 (1258)	617 (620)	306 (237)	121 (69)	26 (14)	0 (0)
England (total)	159	3340 (3156)	1554 (1466)	863 (796)	510 (406)	227 (153)	73 (33)	0 (0)
Wales	12	2284 (1862)	1062 (855)	704 (533)	509 (387)	228 (166)	171 (87)	221 (47)

Table 4 – Estimates of effects of targets

	Num. of patients waiting (Wales not treated)			
	(1) >18months	(2) >15m & <18m	(3) >12m & <15m	(4) >9m & <12m
18mo target	-192.0 (92.6)*			
15mo target		-191.4 (45.0)**		
12mo target			-193.9 (20.5)**	
9mo target				-273.2 (36.1)**
Observations	4682	4682	4682	4682
R-squared	0.76	0.66	0.65	0.76

Robust standard errors in parentheses, * significant at 5%; ** significant at 1%
All regressions include trust, time dummies and a Wales-specific time trend. The standard errors are clustered at trust level.

Table 5 – Estimates of effects of targets (including earlier targets' effects)

	Num. of patients waiting (Wales not treated)						
	(1) >18months	(2) >15m & <18m	(3) >12m & <15m	(4) >9m & <12m	(5) >6m & <9m	(6) >3m & <6m	(7) <3m
18mo target	-192.0 (92.6)*	-55.5 (28.8)	-110.7 (42.3)**	29.1 (48.2)	369.0 (41.2)**	399.7 (54.5)**	300.7 (152.1)*
15mo target		-177.1 (38.7)**	-277.4 (46.9)**	-148.5 (39.0)**	187.2 (31.5)**	302.5 (51.0)**	142.8 (82.3)
12mo target			-217.6 (27.7)**	-252.4 (32.2)**	30.4 (35.4)	108.7 (48.9)*	-14.0 (83.7)
9mo target				-284.9 (44.3)**	-225.3 (53.6)**	91.7 (77.5)	233.2 (102.0)*
Observations	4682	4682	4682	4682	4682	4682	4682
R-squared	0.76	0.66	0.66	0.76	0.84	0.91	0.96

Robust standard errors in parentheses, * significant at 5%; ** significant at 1%
All regressions include trust, time dummies and a Wales-specific time trend. The standard errors are clustered at trust level.

Table 6 – Estimates of effects of targets (Welsh targets considered treatment)

	Num. of patients waiting (Wales targets considered treated for t>24)						
	(1) >18months	(2) >15m & <18m	(3) >12m & <15m	(4) >9m & <12m	(5) >6m & <9m	(6) >3m & <6m	(7) <3m
18mo target	-191.9 (80.4)*	-9.0 (13.5)	17.2 (13.3)	151.1 (36.3)**	305.6 (34.7)**	291.0 (46.7)**	265.0 (148.6)
15mo target		-136.2 (31.5)**	-183.2 (42.9)**	-64.1 (28.4)*	135.9 (21.1)**	183.5 (38.7)**	108.8 (51.0)*
12mo target			-123.9 (15.7)**	-168.0 (19.9)**	-20.9 (24.4)	-10.4 (34.7)	-47.9 (48.8)
9mo target				-119.1 (69.2)	68.6 (69.5)	346.1 (101.0)**	489.0 (177.8)**
Observations	4682	4682	4682	4682	4682	4682	4682
R-squared	0.77	0.68	0.67	0.76	0.84	0.91	0.96

Robust standard errors in parentheses, * significant at 5%; ** significant at 1%
All regressions include trust, time dummies and a Wales-specific time trend. The standard errors are clustered at trust level.

Table 7 – Estimates of effects of targets on mean waiting times

	Mean waiting time (full effects in England without treatment effects in Wales)			
	(1)	(2)	(3)	(4)
18mo target	0.176 (0.133)	0.176 (0.259)	-0.037 (0.165)	-0.037 (0.255)
15mo target	-0.583 (0.123)**	-0.583 (0.149)**	-0.738 (0.156)**	-0.738 (0.194)**
12mo target	-0.731 (0.132)**	-0.731 (0.144)**	-0.886 (0.162)**	-0.886 (0.162)**
9mo target	0.629 (0.126)**	0.629 (0.209)**	0.377 (0.188)*	0.377 (0.193)
Wales t-trend	No	No	Yes	Yes
s.e.	Robust	Cluster(trust)	Robust	Cluster(trust)
Observations	4682	4682	4682	4682
R-squared	0.85	0.85	0.85	0.85

Robust standard errors in parentheses, * significant at 5%; ** significant at 1%
All regressions include trust and time dummies.