Use of Targets to Improve Health System Performance: English NHS Experience and Implications for New Zealand

Nicholas Mays

NEW ZEALAND TREASURY

WORKING PAPER 06/06

JULY 2006

THE TREASURY

Kaitohutohu Kaupapa Rawa

NZ TREASURY WORKING PAPER 06/06

Use of Targets to Improve Health System Performance: English NHS Experience and Implications for New Zealand

MONTH/YEAR

July 2006

AUTHOR

Nicholas Mays

Principal Adviser (part-time), Health Section, Social Policy Branch
The Treasury

and

Professor Of Health Policy
Health Services Research Unit
Department Of Public Health & Policy
London School Of Hygiene & Tropical Medicine
University Of London

Email Nicholas.Mays@lshtm.ac.uk
Telephone +44-20-7927 2222
Fax +44-20-7580 8183

ACKNOWLEDGEMENTS

I am grateful to Professors Julian Le Grand and Gwyn Bevan, London School of Economics and Political Science, and to many colleagues in the New Zealand Treasury and Ministry of Health for helpful comments on an earlier version of this paper.

NZ TREASURY

New Zealand Treasury
PO Box 3724
Wellington 6008
NEW ZEALAND

Email information@treasury.govt.nz
Telephone 64-4-472 2733
Website www.treasury.govt.nz

Abstract

The setting of quantitative, time-limited 'targets' backed up by institutional and managerial rewards and sanctions has been a notable feature of performance improvement efforts in the National Health Service (NHS) in England since 1998 and especially in the period 2000-2004.

Performance improved in the areas covered by English NHS targets, most markedly in relation to waiting times, but also in relation to treatment outcomes. None of the other parts of the United Kingdom followed England and similar trends were not observed, particularly not in waiting times, despite similar injections of funds.

Despite the improvements in performance in target areas, targets were criticised, principally, for having perverse and unintended consequences (e.g. distorting priorities, encouraging 'gaming', etc) which could have potentially out-weighed their benefits. On the other hand most experts in performance improvement in public services argue that carefully chosen, incentivised targets are a useful part of the performance management repertoire when used well (e.g. when sanctions and rewards are proportionate). Some dysfunctional consequences are to be expected, but can be mitigated.

Given the similarities between the English NHS and the New Zealand public health system, there is scope to use targets and related incentives sparingly to improve performance in New Zealand in areas of high importance to government and the public.

JEL CLASSIFICATION I18

KEYWORDS performance measurement; performance management; performance improvement methods; target-setting; control of public services

Table of Contents

List of Tables

List of Figures

Use of Targets to Improve Health System Performance: English NHS Experience and Implications for New Zealand

1 Summary

The setting of quantitative, time-limited 'targets' by government on behalf of patients and tax payers backed up by rewards and sanctions related to performance in meeting these targets has been a notable feature of performance improvement efforts in the National Health Service (NHS) in England since 1998 and especially in the period 2000-2004.

This paper outlines the system of 'star ratings' (global ratings based on performance in relation to a range of targets) and 'league tables' used in the English NHS; summarises the evidence on the impact of this regime; discusses the main criticisms of the use of targets; assesses the strengths and weaknesses of the use of targets to drive performance management; describes what has been learned; and draws out the implications for New Zealand's public health system.

The star rating system and publication of league tables was accompanied by an incentive system that was directed at holding the boards and, especially, the chief executives, of hospitals and other NHS organisations accountable for the local delivery of national priorities. The chief executives of 'zero' rated NHS Trusts were at risk of dismissal along with their chairs. High performing NHS Trusts were able to take advantage of an incentive system that focused on 'earned autonomy' as a reward for success. In addition, funds were set aside at provider level (not controlled by the commissioners/purchasers) to support Trust incentive and reward schemes.

Performance improved over time in the areas covered by English NHS targets, most markedly in relation to waiting times, but also in relation to treatment outcomes. Performance began to improve before most of the recent growth in NHS spending and capacity. The trends suggest that the setting of targets with related league tables and incentives was causally associated with a substantial part of the observed improvement. For example, in 1999, 4.4% of patients on hospital waiting lists had waited more than 12 months before receiving treatment. By 2003, only a handful of patients (less than 0.1%) were waiting more than 12 months. In the period after targets were set, premature deaths from cancers and coronary heart disease fell faster in England than in any other European

country, though admittedly from a high base. For example, the cancer death rate for people under 75 years fell by over 12% between 1999 and 2003.

None of the other parts of the United Kingdom followed England in setting quantitative, time-limited targets, backed up by managerial rewards and sanctions, and similar trends were not observed, particularly not in waiting times, despite similar injections of funds. Indeed, waiting times deteriorated in 1999-2001 outside England. In Scotland, comparative clinical performance data were published, but without the incentives and external scrutiny put in place in England, the initiative had no discernible effect on performance. Progress in England eventually put pressure on other parts of the United Kingdom to emulate England.

Despite the improvements in performance in target areas, targets, and particularly, global star ratings, were criticised, principally, for being too rigid (i.e. not allowing for local variation in priorities), and for having perverse and unintended consequences (e.g. distorting managerial and clinical priorities, omitting important aspects of performance, over-simplifying performance assessment, encouraging 'gaming', etc) which could potentially have out-weighed their benefits. On the other hand most experts in performance improvement in public services argue that carefully chosen, incentivised targets are a useful part of the performance management repertoire when used well (e.g. when sanctions and rewards are proportionate) since they can focus organisations positively on the goals of government. Some dysfunctional consequences are to be expected, but can be mitigated (e.g. by modifying the way in which performance is measured).

The policy challenge is to maximise the benefits and minimise the costs of targets rather than ignoring their potential contribution to performance. Given the similarities between the English NHS and the New Zealand public health system, there is scope to use targets and related incentives to improve performance in New Zealand in areas of high importance to government and the public. On the other hand, any target regime would need to be crafted to take account the particular features of the New Zealand system.

2 Introduction

Performance measurement[1] and management are important elements in the range of methods available to improve performance in publicly financed health systems. There are three aspects: measurement; analysis; and action. Taken together, they are seen as contributing to solving the many principal-agent problems which exist in health care systems due to information asymmetries, and divergent incentives and goals between principals and agents. While performance measurement is widespread in OECD countries and is used, in principle, to improve accountability of agencies in the health sector, raise performance and provide consumers with information about health care choices and quality, analysis leading to action is much less frequently found and its consequences relatively poorly understood.

[1] Schyve (1995, p231 cited in Mannion and Goddard, 2002) defines *performance measurement* as, 'the quantitative measurement of the results of health care (e.g. patient health outcomes, patients' and other customers' satisfaction and resource use), and of processes that are so closely associated with the results of care, that they can be used as surrogates for the anticipated results.' The Joint Commission on Accreditation of Health Care Organisations (1990, p1 cited in Mannion and Goddard, 2002) defines a *performance indicator* as: 'a quantitative measure that can be used to monitor and evaluate the quality of important governance, managerial, clinical and support functions that affect patient outcome.'

One particular approach to performance measurement and management which has attracted considerable attention in the last five years or so, is the setting of explicit, quantitative, time-limited 'targets' by governments and other funders of health services backed up by an incentive regime of rewards and sanctions related to performance in meeting these 'targets'. Such a target-driven performance management regime has been an important component in the wider performance improvement efforts in the National Health Service (NHS) in England especially since the publication of the NHS Plan in 2000 and particularly in the period 2000-2004. From 2004, more emphasis has been given to individual patient choice as a driver of improvement though targets remain in place. Given the similarities between the objectives and challenges facing the publicly financed health systems of New Zealand and the UK, it is worth considering what can be learned from the English experience of using targets to improve performance.

This paper outlines the system of 'star ratings' and 'league tables' used in the English NHS from 1998 to 2005; summarises the evidence on the impact of this regime; discusses the main criticisms of the use of targets; assesses the strengths and weaknesses of the use of targets to drive performance management; describes what has been learned; and draws out the implications for New Zealand's public health system.

In the context of public services and performance improvement, a 'target' can be seen as 'a desired process or outcome that has been codified.' (Collins et al, 2005, p1); i.e. the process or outcome has not only been specified, but it has been incorporated into some system of hierarchical oversight, performance monitoring and accountability. A 'target' is thus a performance indicator embedded in a particular set of organisational incentives.

In the UK context, government 'targets' are quantified, time-specific goals which are built into the Public Service Agreements (PSAs) which have been negotiated annually since 1998 between the central government departments responsible for particular public services and the Treasury on the basis of which funding is made available, and which are signed by the relevant ministers and the Chancellor of the Exchequer (see Box 1).

Box 1: UK Public Service Agreements (PSAs)

Public Service Agreements (PSAs) were introduced in the 1998 Comprehensive Spending Review (CSR). They set out agreed targets detailing the outputs and outcomes departments are expected to deliver with the resources allocated to them. The new spending regime places a strong emphasis on outcome targets, for example in providing for better health and higher educational standards or service standards. The Government monitors progress against PSA targets, and departments report in detail twice a year in their annual Departmental Reports (published in spring) and in their autumn performance reports. These reports provide Parliament and the public with regular updates on departments' performance against their targets. Technical Notes explain how performance against each PSA target will be measured.

In the case of the English NHS, the Department of Health's PSA 'targets' (Box 2 sets out the Department's current targets following the 2004 Spending Review) are incorporated into the managerial performance assessment and incentive system operated by the Department of Health for NHS provider organisations and commissioners (Primary Care Trusts (PCTs)).

Box 2: Department of Health PSA Objectives and performance targets, 2005-2008

Objective I: Improve the health of the population. By 2010 increase life expectancy at birth in England to 78.6 years for men and to 82.5 years for women

1. Substantially reduce mortality rates by 2010:

- *from heart disease and stroke and related diseases by at least 40% in people under 75,with at least a 40% reduction in the inequalities gap between the fifth of areas with the worst health and deprivation indicators and the population as a whole;*
- *from cancer by at least 20% in people under 75,with a reduction in the inequalities gap of at least 6% between the fifth of areas with the worst health and deprivation indicators and the population as a whole; and*
- *from suicide and undetermined injury by at least 20%.*

2. Reduce health inequalities by 10% by 2010 as measured by infant mortality and life expectancy at birth.

3. Tackle the underlying determinants of ill health and health inequalities by:

- *reducing adult smoking rates to 21% or less by 2010, with a reduction in prevalence among routine and manual groups to 26% or less;*
- *halting the year-on-year rise in obesity among children under 11 by 2010 in the context of a broader strategy to tackle obesity in the population as a whole.* Joint with the Department for Education and Skills and the Department for Culture, Media and Sport; *and*
- *reducing the under-18 conception rate by 50% by 2010 as part of a broader strategy to improve sexual health.* Joint with the Department for Education and Skills.

Objective II: Improve health outcomes for people with long-term conditions

4. Improve health outcomes for people with long-term conditions by offering a personalised care plan for vulnerable people most at risk; and to reduce emergency bed days by 5% by 2008, through improved care in primary care and community settings for people with long-term conditions.

Objective III: Improve access to services

5. Ensure that by 2008 no-one waits more than 18 weeks from GP referral to hospital treatment.

6. Increase the participation of problem drug users in drug treatment programmes by 100% by 2008 and increase year on year the proportion of users successfully sustaining or completing treatment programmes.

Objective IV: Improve the patient and user experience

7. Secure sustained national improvements in NHS patient experience by 2008, as measured by independently validated surveys, ensuring that individuals are fully involved in decisions about their healthcare, including choice of provider.

8. Improve the quality of life and independence of vulnerable older people by supporting them to live in their own homes where possible by:

- *increasing the proportion of older people being supported to live in their own home by 1% annually in 2007 and 2008; and*
- *increasing by 2008, the proportion of those supported intensively to live at home to 34% of the total of those being supported at home or in residential care.*

3 Policy initiatives and systems using targets to improve performance in the English NHS

The most widely discussed, even notorious part of the English NHS system of performance improvement derived from the PSAs and related targets, were the so called '*Star ratings'* in which both provider organisations such as hospitals and mental health providers, and commissioning bodies (i.e. PCTs) were given an overall rating, initially by the Department of Health and after 2004 by an independent inspectorate, from zero to three stars based on their performance on a number of target measures. Although NHS organisations were subject to a regime of performance measurement which included approximately 40 indicators arranged according to six dimensions of performance[2], in practice, the star ratings for providers, 2000-2004 depended on performance on nine 'core' indicators mostly related to waiting times[3] and financial stability which were regarded as largely within their control. PCTs were assessed according to the same waiting times targets as NHS provider organisations plus targets relating to time to see a GP, time to see another primary care professional, time to access drug misuse treatment and 4-week smoking quit rates. Annual star ratings in the form of so called 'league tables' were published from 2001 to 2005.

The policy goal of the 'star ratings' system was to provide the mix of managerial and financial incentives to performance improvement in areas of high priority to government and to patients which the internal market of the 1990s was regarded as having failed to produce. They were also a response to a perceived deterioration of performance in relation to waiting times since the mid-1990s together with a number of unfavourable comparisons of the outcomes of care in England with continental Europe[4]. The decision to compress a range of performance data into a single rating was a deliberate attempt to summarise complex information in the form of rankings that the public and the media could relate to, thereby promoting public interest, strengthening public accountability, and encouraging managerial and clinical action. It was envisaged that neither managers nor clinicians would wish to be associated with poorly performing institutions once this was publicly known and would respond by trying to improve their relative position.

The rating system was accompanied by an incentive system that was directed at holding the boards and, especially, the chief executives of hospitals and other organisations accountable for the local delivery of national priorities through the 'naming and shaming' which inevitably accompanied the publication of league tables[5]. Given that there was a purchaser-provider split in the English NHS, there was always the possibility that purchasers might attempt to redirect some of the work of providers in response to poor

[2] The dimensions of the Performance Assessment Framework (PAF) are: population health improvement; fair access; effective delivery of appropriate care; efficiency; patient and carer experience; and health outcomes after treatment.

[3] There were three waiting times targets that were given particular weight:

- Percentage of A&E patients seen within 4 hours of attendance (2003-05 target was 90% and 98% in 2005);
- From 2002 75% of category A ambulance calls were to be responded to within 8 minutes;
- Maximum waiting time for first elective hospital admission of 18 months in 2001, 15 months in 2002, 12 months in 2003, 9 months in 2004 and 6 months in 2005 (3 months for outpatients).

The current target is that by 2008, no one is to wait more than 18 weeks from GP referral to hospital treatment, less for most patients, and quicker still for cancers.

[4] For example, a review of cancer services followed the reporting of poorer survival of women in the NHS following treatment for breast cancer than in parts of continental Europe such as France and Germany. Among other things, the review led to the setting of a target maximum two-week wait for referral from the GP to a cancer specialist to identify treatable cases sooner.

[5] NHS Trust boards comprise both executive directors (full-time senior managers including the chief executive and director of finance) and non-executives (part-time, ministerial appointees including the chair), thereby enabling the centre to hold senior managers directly accountable for performance alongside other members of the governance board.

performance ratings. The chief executives of zero rated NHS Trusts were also at risk of dismissal along with their chairs. Some of the remuneration of senior staff was potentially at risk if performance was poor. On the other hand, high performing NHS Trusts were able to take advantage of an incentive system that focused on 'earned autonomy' as a reward for success. In addition, £155m was set aside in 2001 as a performance fund at provider level (not controlled by the commissioners/purchasers) to support Trust incentive and reward schemes. The better the performance, the more autonomy Trusts were given in terms of how to spend the extra funds as well as a lower level of oversight from the centre. Funds were to be spent on and by the staff who had contributed to meeting the performance targets. In addition, three-star Trusts were able to apply for 'Foundation' status in the first wave of this new, more autonomous form of governance of NHS provider organisations. Finally, three-star Trust management teams had the opportunity to take over the running of 'failing' Trusts thereby sharing their experience and expertise more widely and building their reputations. Although the system of reducing more complex assessments of performance to a single 'star' rating has now been abandoned in favour of a more complex, multi-dimensional external assessment of performance by the quality regulator, the Health Care Commission, the use of 'targets' remains a central part of the performance improvement system in the NHS in England, now focusing increasingly on reducing waiting time before initial diagnosis rather than from diagnosis to treatment.

4 The impact of the 'target' regime in the English NHS

In general, performance improved markedly in the areas covered by English NHS 'targets'. The timing of these improvements, and the fact that they did not occur elsewhere in the UK, suggest that the setting of targets with related league tables and incentives was causally associated with a substantial part of the improvement.

4.1 Waiting times

Table 1 summarises the trends in the English waiting time targets first set in 2000 and given high priority by Ministers and the Department of Health through the performance management regime. It shows that on all the indicators performance improved markedly comparing the situation before the target was introduced with the years that followed. It is noteworthy that performance had already begun to improve by 2002/03 before most of the recent growth in NHS spending and capacity began, and that similar trends are not observed in Scotland, Wales and Northern Ireland despite similar injections of funds (see below).

The effect on performance of setting and focusing management effort on the A&E and ambulance key targets for England are shown graphically in Figures 1 and 2. Note the average improvement and the narrowing of the gap between the poorest and best performing ambulance services in Figure 2.

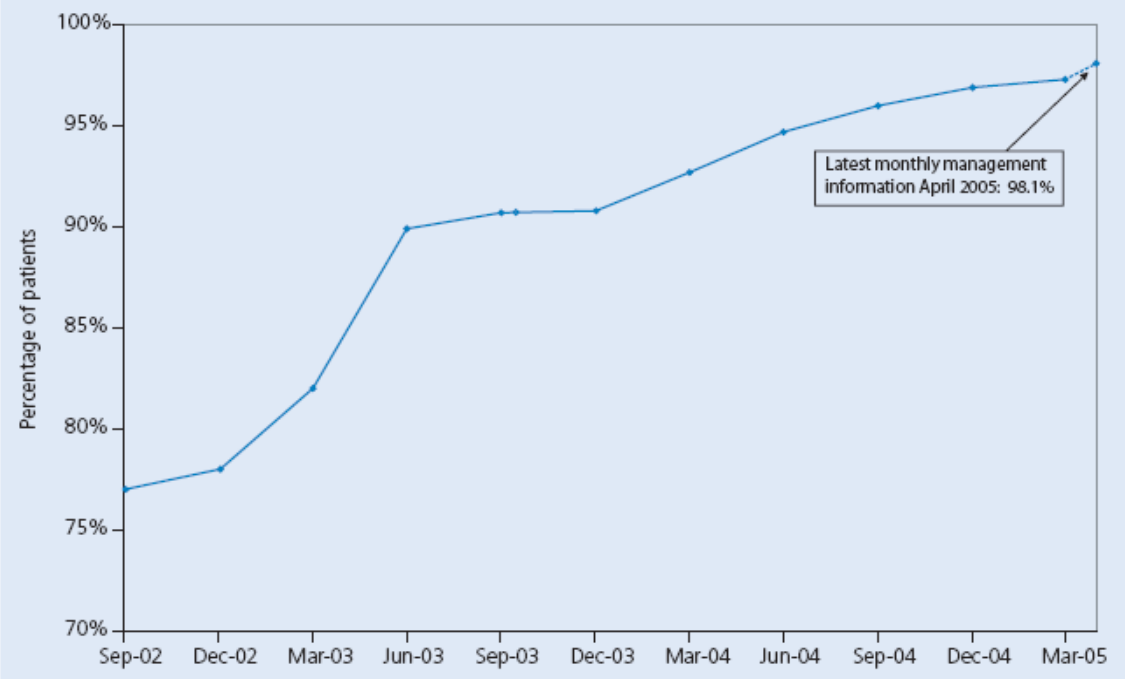
Figure 3 shows the downward trend in patients waiting more than six months for inpatient admission. No one waited more than 12 months by March 2003. Other performance data relating to waiting are shown in Appendix 1.

Table 1: Effects of the English NHS targets for waiting, 1999-2005

Target	*1999*	*2000 (targets set)*	*2001 (first 'star ratings')*	*2002*	*2003*	*2004*	*2005*
90% A&E patients <4 hrs, 2003-05 & 98%, 2005				23% waiting >4 hours		5.3% waiting >4 hours	
75% ambulance category A in ≤8 mins		1/17 Trusts met target					14/17 met target
Max wait for first elective admission: 18m, 2001 15m, 2002 12m, 2003 9m, 2004		67,000 patients waiting >12m 185,000 waiting >9m					24 >12m 41 >9m 41,000 >6m
% on list waiting >12m	4.4	4.7	4.2	2.1	0	0	0
% on list waiting >6m	26.1	25.8	24.4	23.3	19.4	8.9	5.0

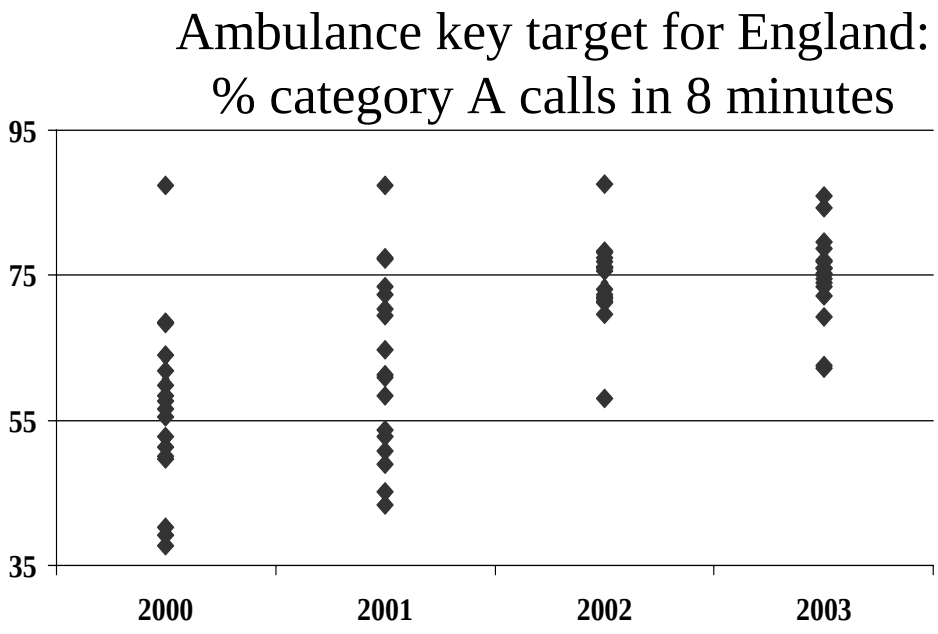
Source: Department of Health

Figure 1: Percentage of patients spending four hours or less in A&E



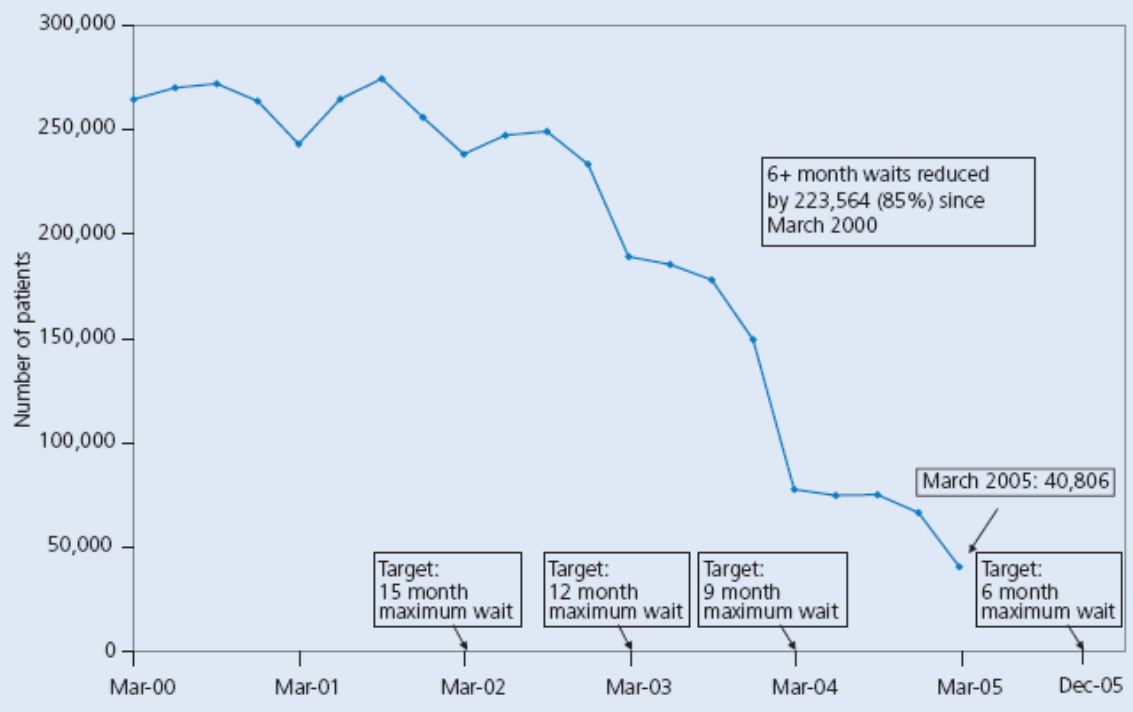
Source: Department of Health (2005) *Chief Executive's Report to the NHS*. London: Department of Health

Figure 2



Source: Gwyn Bevan, London School of Economics and Political Science, analysis of Department of Health data

Figure 3: Patients waiting over six months for admission



Source: Department of Health (2005) *Chief Executive's Report to the NHS*. London: Department of Health

4.2 Comparison between England and the rest of the UK

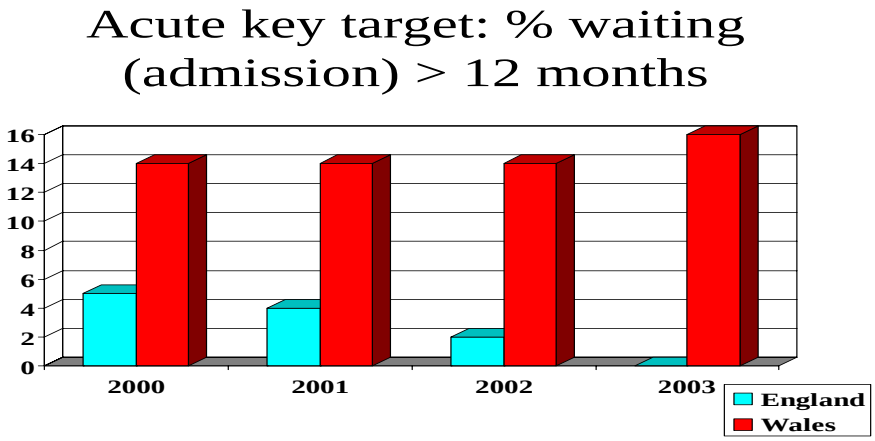
None of the other parts of the United Kingdom followed England in setting high profile, quantitative, time-limited targets, monitoring individual indicators, publishing overall 'league tables' and operating an incentive regime of organisational and managerial rewards and sanctions applying both to commissioners and providers of services. Indeed, after political responsibility for the NHS was devolved to Wales and Scotland in 1998, targets were abandoned. For example, in Wales, while the separation between purchase and provision remained, waiting time targets were removed and policy focused instead on cooperative working between health, local government and the voluntary sector to improve community health. In Scotland, the 1990s quasi-market was abolished, there were no English-style 'targets' and steps taken to build a professionally led, integrated system based on concepts such as the managed clinical network. As a result, there has been a natural experiment into the effect of 'targets' in the context of a quasi-market (England) versus no 'targets' in the context of a more collaborative, network-based system (Scotland and Wales). In fact, England also came to experiment with clinically led networks (so called 'collaboratives') in order to improve performance in particular areas such as cancer (see below). Thus the English approach was a hybrid with elements of 'hierarchical', 'market' and 'network' governance.

The most striking difference in the performance of the NHS across the four countries of the UK comparing 1996 with 2003 (when 'targets' dominated in England) was the reported reductions in waiting in England which did not occur elsewhere despite similar funding increases in the other countries of the UK and which seemed to be the result of strong performance management against targets (Alvarez-Roseté, Bevan, Mays and Dixon, 2005). The bar chart below (Figure 4) shows the percentage waiting more than 12

months in England and Wales between 2000 and 2003. The percentage rose in Wales while waits of more than a year virtually disappeared in England.

Hauck and Street (2004) compared the performance of hospitals on either side of the England-Wales border over a six-year period before and after devolution. They showed that the English hospitals recorded increased levels of activity, undertook proportionately more day case activity and had declining mortality rates while activity levels remained constant in Wales, the proportion of day cases fell and mortality rates rose. English patients waited less time and were more likely to be treated within the target waiting period.

Figure 4



Source: Gwyn Bevan, London School of Economics and Political Science

Table 2 gives the comparative trends for 6-month and 12-months waits in England, Wales and Northern Ireland (unfortunately, Scottish waiting time statistics are not comparable with the rest of the UK). It shows deterioration in Wales and Northern Ireland, 1999-2001, when England's performance was improving markedly followed by some signs of a catch-up as pressure mounted to emulate trends in England. For example, latterly the NHS in Wales has given much more attention to reducing long waits with considerable success.

Table 2: Percentage of patients on NHS hospital waiting lists waiting longer than 6 or 12 months, 1999-2005

	1999	2000	2001	2002	2003	2004	2005
% waiting >12 months							
England	4.4	4.7	4.2	2.1	0	0	0
Wales	11.2	14.2	13.8	14.3	15.9	11.3	1.3
Northern Ireland	17.9	20.0	21.8	24.9	22.0	14.7	8.5
% waiting >6 months							
England	26.1	25.8	24.4	23.3	19.4	8.9	5.0
Wales	NA	NA	34.0	37.0	37.0	35.2	24.9
Northern Ireland	36.7	39.1	41.4	44.1	40.0	34.1	28.1

NA=data not available.

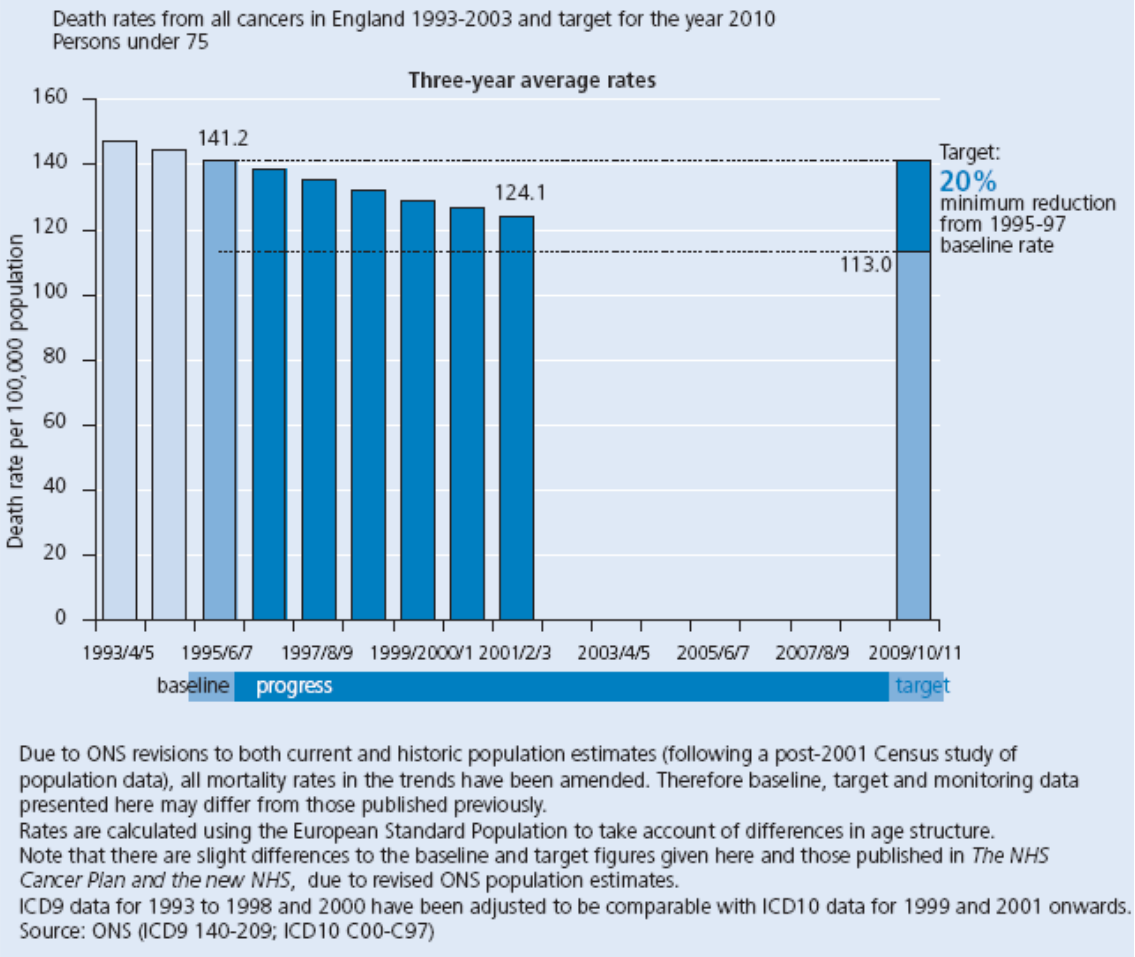
Source: Bevan G, Hood C. Have targets improved performance in the English NHS? *BMJ* 2006; 332: 419-22

4.3 Trends in treatment outcomes

Targets were set in England to reduce death rates from cancers, circulatory disease and intentional self-harm from the levels when the Labour government took power in 1997. Although analysis of trends cannot prove causality, the focus on these three areas appears to have been associated with noticeable improvements in mortality (see Figures 5-7). In the period after targets were set, premature deaths from cancers and coronary heart disease fell faster in England than in any other European country, though admittedly from a high base. For example, the cancer death rate for people under 75 years fell by over 12% between 1999 and 2003. The cancer target has been associated with a range of efforts to map and reorganise the patient pathway through the cancer system both at a regional level and within individual parts of the cancer service (e.g. speeding up non-urgent waits for screening by setting up clinics led by specialist nurses or nurse-led endoscopy clinics for urgent referrals). This so called 'modernisation' of service delivery and organisation was given greater urgency by the existence of clinical targets. There was also a large increase in the number of cancer specialists (approximately 38% between 1997 and 2005) precipitated by the requirement to meet the cancer target.

In the case of coronary heart disease there was a 43% increase in revascularisations with much shorter waits from referral to treatment and a 227% increase in the prescription of lipid lowering medications between the publication of the NHS Plan in 1999/2000 and 2004/05. In the mental health field, the suicide rate fell 6% between the baseline rate (1995-97 average) and 2001-03 average (see Figure 7). This was accompanied by a large increase in spending, staff and access to new treatments, but, again, the initial significant improvements predated the large increases in the NHS budget from 2002/03.

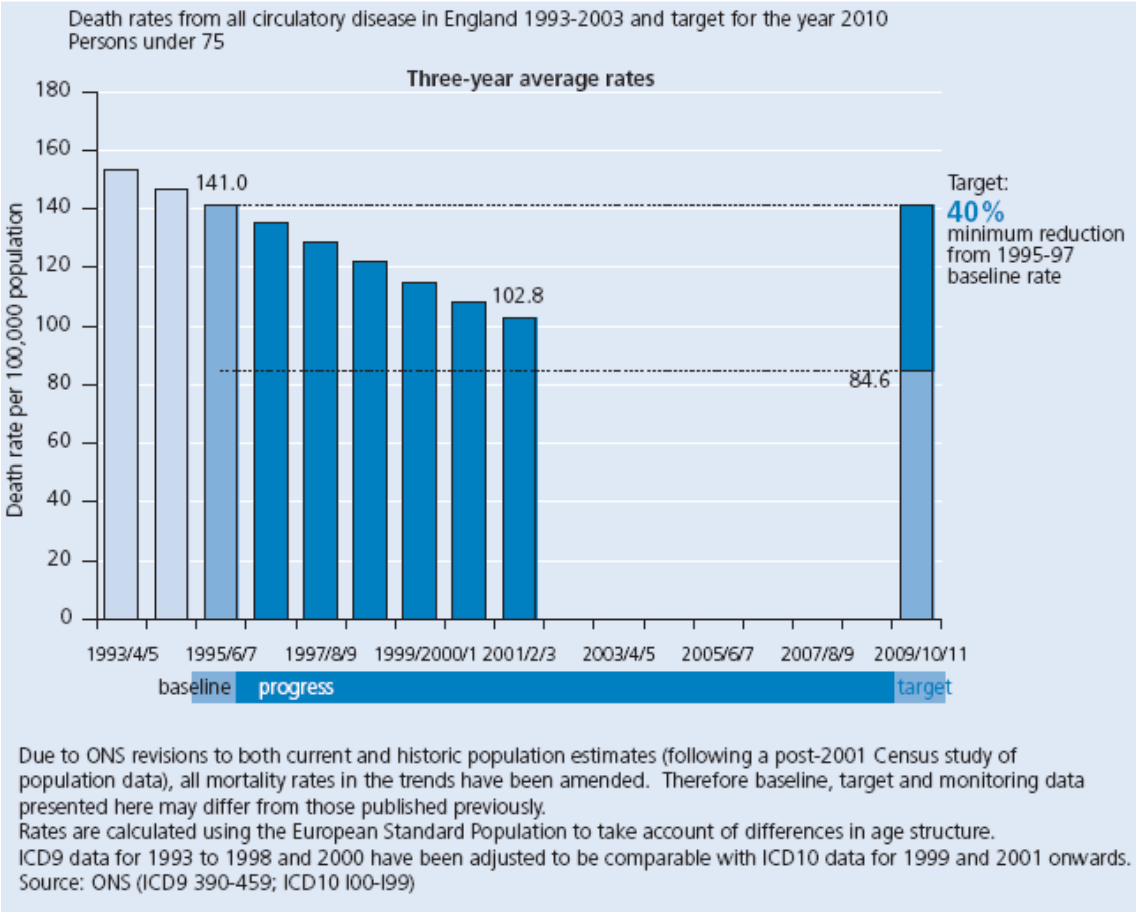
Figure 5: Progress against cancer mortality target



Source: Department of Health (2005) *Chief Executive's Report to the NHS*. London: Department of Health

In Scotland, clinical performance data were published in the 1990s, but without external quality assurance of the data and without the incentives and external scrutiny put in place in England. It was expected that providers and clinicians would use the data to stimulate further investigation into causes of performance variation and act accordingly. The initiative had no discernible effect on performance (Mannion and Goddard, 2001).

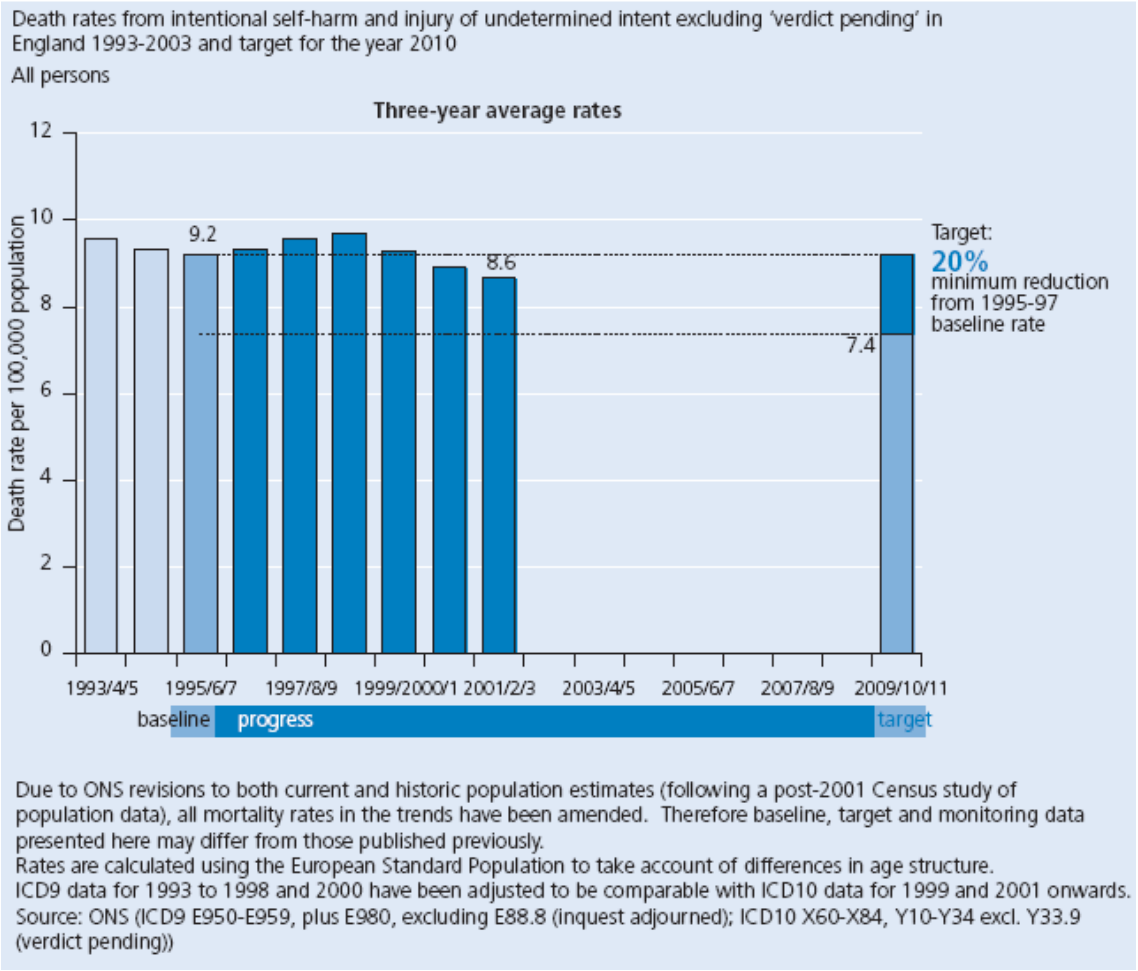
Figure 6: Progress against circulatory disease mortality target



Source: Department of Health (2005) *Chief Executive's Report to the NHS*. London: Department of Health

The markedly superior performance in England on those indicators which were the subject of English 'targets' did not mean that England was the best performer on all the aspects of performance which it is possible to measure comparably across the UK. For example, Scotland had the highest rates of breast cancer screening coverage of women aged 50-64 (2002), influenza vaccination for over 65s (2004) and MMR vaccination coverage of 2 year olds (2001) of the four countries. Wales had the highest rates of statin prescribing. However, England had the lowest all-cause mortality, and specifically, the lowest mortality from colorectal cancer and coronary heart disease.

Figure 7: Progress against mental health target



Source: Department of Health (2005) *Chief Executive's Report to the NHS*. London: Department of Health

In addition, there was no obvious sign in the non-target areas that performance in England was being harmed by efforts to raise performance in target areas despite the fact that per capita expenditure was substantially lower in England than in other parts of the UK (Alvarez-Roseté, Bevan, Mays and Dixon, 2005). It is, however, possible that the policies and approaches to performance improvement of the other parts of the UK may take longer to produce their benefits and that these gains could become apparent in the future.

5 Criticisms of the use of targets and star ratings in the English NHS

Despite the improvements in performance in areas which were the subject of targets and related incentives, which was not matched elsewhere in the UK, targets, and particularly, star ratings, have been criticised, principally for being crude and leading to perverse consequences. Many of the criticisms, though expressed in a variety of ways, amount to an argument against the priorities which the targets represented. Some of the criticisms related to the specific way in which the achievement of targets was used to produce overall star ratings together with the sanctions and rewards that followed different ratings. Other criticisms were intrinsic to target regimes in general.

5.1 There were too many targets

It is undoubtedly true that there were a large number of targets potentially contributing to the star rating system. Efforts have been made since the early 2000s to reduce their number. It was also apparent that a much smaller number of targets, mostly related to the responsiveness of services, dominated the eventual calculation of star ratings. This criticism is not intrinsic to the use of targets, though there is a general tendency for the number of targets to grow over time unless carefully managed.

5.2 Targets were too rigid and undermined staff morale

These criticisms related to the fact that targets did not take into account local variations in factors such as the incidence of disease (size of the problem), deprivation (difficulty of implementing a response) or the fact that different places may have different local problems. This was principally because the most important targets initially related to process improvements (i.e. shorter waits) which were regarded by the Department of Health as directly under the control of the NHS irrespective population characteristics. The concern about staff morale relates to the general risk that too much emphasis on externally driven targets may drive out intrinsic motivation, particularly among professional staff. This reinforces the point above about keeping the number of targets under control.

This set of criticisms has to be taken seriously. It challenges those setting targets to set realistic targets in areas that local health services can influence, consult front-line staff and service users on these, and be prepared to refine targets in light of experience. However, in tax-financed systems, final authority (and responsibility, arguably) for setting targets should rest with central government, despite the fact that it is tempting to allow influential professionals to set them (Collins, Sibson, Shotas, Smith and Thornton, 2005). While local managers and clinicians will need more detailed performance and other information for running services, this should be seen as quite distinct from national targets.

5.3 Targets had perverse and unintended consequences

This set of criticisms has been widely debated and is true in the sense that participants in systems may, depending on motivation, try to find ways of minimising their need to change in order to meet performance improvement targets. There was evidence of perverse effects in the English NHS case (e.g. GPs' receptionists refusing to allow patients to book appointments more than 48 hours ahead, even when it suited them, on the basis that this would count against the practice hitting the maximum 48-hour wait target. The basis of measurement was subsequently altered to exclude appointments more than 48 hours ahead where the patient had requested this). There were reports of hospitals reducing the number of patients on waiting lists and driving down average waits by preferentially treating 'easier' patients at the expense of more complex patients who might have had to wait longer as a result (Chang, 2006).

The force of this set of criticisms can be reduced with greater transparency about why a particular target had been set and measuring performance in ways that relate to the underlying rationale for the target. As Collins et al (2005) point out, there is no intrinsic reason why a target should *distort* priorities since, by definition, a target *indicates* a priority. Seen from this perspective, many criticisms of the 'distorting' effects of targets are, in fact, criticisms of the priorities themselves and of the fact that they focused

managerial and clinical attention on particular areas (e.g. cancer treatment) and/or facets of performance (e.g. waiting) rather than others. For instance, in England the government's school students' literacy targets were criticised for 'narrowing' the curriculum and driving out other educational activities, but if literacy cannot be taught acceptably within a broad curriculum, arguably the curriculum has to be narrowed as long as literacy is regarded as an important outcome of schooling. In the NHS, there was criticism that excessive emphasis on reducing waiting times was leading hospital departments to treat patients according to the length of time they had waited irrespective of clinical priority or, as mentioned above, in relation to ease of treatment, and that some patients' health was suffering accordingly. Again, this was a criticism of the relative priority given to average waiting time over other measures of performance in the elective area such as the ability to prioritise patients on the waiting list.

5.4 Failures of performance outside target areas were treated as if they did not matter

This criticism is a specific manifestation of the general concern about distortion and unanticipated, undesirable consequences. It focuses on the signals sent out if some areas are the subject of targets and others not (e.g. if one area is subject of a target because it is measurable and another not because it is intangible, but is arguably of equal value), and the fact that under the 'star rating' system it was possible for a provider to harm performance in one area in order to 'hit' a target since the star rating was an average of performance across a range of indicators. While it is inevitable that targets will indicate that certain aspects of performance are of higher priority than others, they do not have to imply that non-target areas of performance are of no importance. A target regime should be accompanied by monitoring of performance across a far wider range of areas.

5.5 It was not always clear who was responsible for meeting targets

This criticism particularly relates to targets which apply to more than one department or agency. It is now generally accepted that one department or agency should be required to take lead responsibility for delivery of a joint target, perhaps by being given an overall budget for its delivery and contracting other agencies to help so that the responsibility can be sub-divided when drawing up implementation plans. This was not an especially important criticism of the NHS target regime since the initial targets related in the main, though not exclusively, to activities under the potential control of different NHS bodies (e.g. waiting times for specific types of services such as A&E). On the other hand, other targets, such as the target relating to reducing deaths from heart disease and stroke (see Box 2), apply to both primary care and hospital providers.

5.6 Targets did not always take into account factors affecting performance outside the control of NHS organisations

A range of variables may influence measured levels of performance outside the control of the assessed organisation such as differences in the characteristics of the population served, geography and random variations. Other constraints may be controllable, but only

in the longer term, such as the capital stock of a hospital. Funding formulae such as those used in New Zealand and England attempt to compensate health entities for environmental circumstances, but are inevitably imperfect. There is evidence that the star ratings and key performance targets set for English PCTs and providers did not allow adequately for different environmental circumstances, suggesting caution in using targets without adjustment (Jacobs, Martin, Goddard, Gravelle and Smith, 2006). This is another argument for exercising judgement when using the results achieved by organisations to determine rewards and sanctions.

5.7 The data on which targets were based were not credible

Collins et al (2005) argue that this is the biggest single problem with targets in the UK public service context and relates to the measurability of targets, the quality of the data used and the vulnerability of the data to manipulation (another perverse consequence). Measurability per se was less of a problem in the English NHS than it might have been since, not surprisingly, the initial set of highly incentivised targets tended to be set in areas where established data systems were available (to the extent that the government was also criticised for focusing excessively on those aspects of NHS performance that were easily measurable at the expensive of other, potentially equally important aspects of performance!). However, this does not exclude the possibility that a part of the improvement observed was simply a result of changes in the way that data were recorded and reported (e.g. that the reductions in four-hour waits in A&E were largely artefacts of finding 'smart' ways of diverting patients from the A&E department or that improvements in waiting times for elective surgery were the result of 'culling' low priority patients from waiting lists or not accepting as many patients onto the waiting list in the first place). It seems at least possible that part of the reported improvement in waiting times, if not in survival, was due to changes in reporting which did not occur in Scotland, Wales and Northern Ireland because the pressure to reach targets was absent. However, other pieces of evidence indicate that the improvements reported were also genuine. For instance, numbers of elective operations carried out increased markedly in the period. In addition, there was no systematic approach to determining waiting list priorities (unlike in New Zealand under the surgical booking system) in the English system (i.e. all patients on the waiting list were, in theory, regarded as being equally worth treating since the problem of managing elective treatment was defined by government as a problem of delay and insufficient capacity rather than one of inadequate prioritisation), thereby reducing the scope for using prioritisation to remove patients from waiting lists.

One overarching response to this group of criticisms is to put in place an independent agency, as in Canada, to supervise measurement issues (e.g. to prevent accusations of government or individual organisations 'cooking the books'), to develop better data collection systems and to compile performance reports.

5.8 The degree of sanction or reward was not clearly related to the degree of failure or success

It was argued that organisations could be assigned a zero star rating with major consequences through missing just one target even if their performance in other respects was exemplary and that the system allowed no discretion to take this into account. Clearly any system in which performance against targets is strongly incentivised needs to

allow some discretion. Sanctions and rewards need to be proportionate to the degree and extent of any failures or success, not simply to a particular score.

5.9 The 'centre' could not adequately determine what mattered most

This criticism relates to the criticism of rigidity, above, and is based on the observation that local circumstances vary, what matters most will vary and that the 'centre' cannot 'know' sufficiently well to determine a set of national targets that apply always and everywhere. For instance, given the differences in social and ethnic composition of the population in different parts of the country, it was possible that the most important causes of avoidable death or morbidity were sufficiently different that a national target to reduce cancer death rates was far less pressing and appropriate in some parts of the country than in others. Again, this criticism highlights the importance of restricting the number of targets, confining them to major performance issues and/or focusing on higher level outcomes. However, in nationally funded and accountable systems such as tax financed public health systems, central government has a right and a duty to set accountable targets for peripheral agencies which may not be uniformly accepted by these agencies. This does not preclude the setting of local, accountable targets (see below for more on this). Another response to this criticism is to set, monitor and assess performance against targets in terms of *changes* in indicators rather than solely on the attainment of an absolute level of performance.

5.10 The 'gaming' and other negative responses could well have overridden the benefits of having accountable targets

This criticism is ultimately empirical, though difficult to verify[6]. It is conceivable that any neglect of non-targeted areas could have reduced service quality sufficiently to offset any benefits accruing in the targeted areas. In addition, there were reports of perverse behaviour (e.g. refusing to deal with low risk patients at A&E departments at peak times until the four-hour waiting target could be met), responses which had no obvious health care benefit but helped with meeting a target (e.g. employing nurses simply to greet patients at A&E departments to ensure that all arriving patients were 'seen' within five minutes of arrival) and fraud (e.g. manipulation of waiting lists and waiting time statistics). On the other hand, in the English NHS, targets were modified in light of evidence of inappropriate responses to try to mitigate their impact. Thus the target that all patients should be able to book an appointment with their GP within 48 hours was amended to reflect the fact that some patients wished to book appointments further in the future but were being prevented from doing so by practices anxious to hit the target because of the financial consequences. One way to mitigate gaming and other dysfunctional behaviours is to have performance data independently audited (see above). Despite the problems it generated, experts on performance management tend to conclude that the target regime was a net improvement over the previous arrangements in the NHS (see below).

[6] Many of the reports of adverse effects were anecdotal, but some have been credibly documented, for instance, by the House of Commons Public Administration Select Committee (2003)

6 Strengths and weaknesses of the system of 'targets' and 'star ratings'

Like any policy instrument, the preceding discussion and analysis shows that the use of incentivised targets has both strengths and weaknesses. In this sense targets are no different from the other policy instruments available (i.e. variants on exhortation, 'choice' (markets), 'voice' (community governance) and 'targets'). Table 3, below, summarises the strengths and weaknesses of the English experience with 'targets'.

Table 3: Strengths and weaknesses of 'targets' and 'star ratings'

Strengths	*Weaknesses*
Relatively simple, clear targets with strong financial incentives improved performance, especially in relation to waiting times. These gains were not seen in other parts of the UK where specific targets were not set and where performance continued to deteriorate	Managers focused on targets contributing to 'star ratings' rather than on areas that might have been more important for overall health system performance (e.g. a focus on meeting the target of a 2-week maximum wait for suspected cancers risked crowding out follow-up visits)
Progress in England eventually put pressure on other parts of the UK to emulate England	
Targets covered aspects of performance that were of great importance to patients (e.g. trolley waits in A&E for inpatient admission)	Targets potentially neglected important areas since they were limited by available data to what could be measured easily (e.g. a relative paucity of data on clinical quality versus waiting times)
Some targets related to quality of care (e.g. emergency re-admissions) and not just to issues of responsiveness	
It was possible to mitigate expected dysfunctional consequences (e.g. by modifying targets)	Targets produced a range of dysfunctional responses
	Targets and related incentives risked crowding out intrinsic motivation to 'do well'

Advising caution in the use of targets and performance indicators, Carter, Klein and Day (1995) point out that most performance indicators are just that – indicators – rather than definitive measures of performance. They argue that most performance indicators are 'tin openers' rather than 'dials' – that is, they do not give definitive answers to questions of performance, but prompt further investigation. By themselves, such indicators inevitably provide an incomplete and inaccurate picture of performance as a whole. Yet the English NHS targets were never designed to cover a representative range of areas of performance. Instead, there is little doubt that the adoption and pursuit of a small number of targets in the English NHS improved performance in the areas of focus which were regarded politically as being of over-riding importance. Nevertheless, in order to hit some targets, systems may have to be restructured in ways that brought benefits beyond the specific target (e.g. anecdotally, in order to reduce waiting times in A&E departments, it was necessary to change the way in which entire departments operated with attendant gains in the quality of care). Rather than prompting the abandonment of performance

indicators and targets, Carter et al's observations underline the importance of careful selection of target areas while maintaining a system of monitoring of other areas of performance.

Sheila Leatherman, a leading quality improvement expert with experience of both the UK and the US, argues, on the basis of a review of trends in performance in the two countries from the late 1990s, that the 'First Fallacy' of performance improvement in health systems is to claim that performance targets and indicators do not work (Leatherman, 2005). Le Grand (2004) agrees, but argues that the sort of target and incentive regime adopted in the English NHS was generally only likely to be effective in meeting relatively unequivocal performance goals which could be adequately measured by simple indicators (e.g. waiting times). He also argues that such a regime was only likely to work well over a relatively short period of time. As time wears on, the odds of gaming responses and harm to morale would be likely to increase. This suggests that strongly incentivised targets should probably be used sparingly as part of a range of different instruments to improve performance. Le Grand's concerns about the long-term viability of targets also relate to his preference for strengthening user choice and competition as drivers of performance improvement in the English NHS over the use of hierarchical managerial targets.

Despite the acknowledged weaknesses of the target regime, the overall verdict of analysts has tended to be positive. For example, Bevan and Hood (2006, p421) conclude their review of the impact of NHS targets as follows:

> 'Nobody would want to return to the NHS performance before the introduction of targets, with over 20% of patients spending more than four hours in accident and emergency and patients waiting more than 18 months for elective admission. And attempts to improve performance without a star system in Wales were criticised by the auditor general for Wales for having "provided neither strong incentives nor sanctions to improve waiting time performance" and were widely perceived to have rewarded organisations that failed to deliver on waiting times.'

This broad view is shared by the members of a recent independent commission into the use of targets in the public services:

> 'Targets can and should be defended. They ought not, in our view, to be abandoned. They are useful for a government to ensure that public money is spent well. Used well, targets provide organisational focus, embody the ambitions of government and offer a transparent account by which services can be measured.'
>
> (Collins et al, 2005, p1)

From this perspective, the policy challenge is thus how to maximise the social benefits and minimise the costs of a regime of targets and incentives rather than abandoning the entire approach.

7 Learning from the English NHS experience with targets

Drawing together recent commentary and analysis of the UK experience with the use of targets linked to managerial incentives in public services in general and the English NHS more specifically (e.g. Collins et al, 2005), a number of lessons can be distilled with relevance to New Zealand, as follows:

1 Targets with linked incentives add most value where other mechanisms such as user choice ('exit' or the threat of 'exit'), competition between providers for purchasers' contracts, or 'voice' are likely to be infeasible or blunt tools[7].

2 Targets and incentives, as in the English NHS, are likely to be most effective in meeting relatively unequivocal performance goals which can be adequately measured by simple indicators (e.g. waiting times) (Le Grand, 2004). Thus targets should be easily measurable and specify an effect over a defined period of time; i.e. an x% reduction in y over time period y. Targets should allow for uncertainty (i.e. random variation). Both outcome and process targets can be valuable as long as in the case of the latter there is a strong link between change in the process and the likelihood of achieving a desired policy outcome. If measurement is intractable, then targets are unlikely to be the correct instrument for performance improvement and, instead, performance should be monitored using a wide range of qualitative and quantitative evidence to provide warning of inadequate levels of performance.

3 Both outcome and process targets can be valuable as long as, in the case of the latter, there is a strong link between change in the process and the likelihood of achieving a desired policy outcome.

4 Targets should be few in number because they should only be set for aspects of services that are exceptionally important *and* where there is a consensus that change is needed[8]. Targets are best used where there is evidence of service failure, unacceptable levels of performance, new higher performance standards or a consensus that a service needs to be delivered in a radically new way[9].

5 Targets should not be used to maintain the status quo (e.g. after an initial target has been attained), but should be seen as *dynamic* (i.e. once they have been reached, they should be removed and replaced with monitoring). Le Grand (2004) also argues that targets are likely to work best in an area over relatively short period of time. As time wears on, the odds of 'gaming' responses and harm to morale are likely to increase. This suggests that strongly incentivised targets should probably be used sparingly for a specific period as part of a range of different instruments to improve performance.

[7] Targets were used in the English NHS alongside a range of different tools such as collaboratives, regionalisation of services and investment in increased capacity, and, latterly, a range of market-related instruments such as patient choice and public-private competition to deliver services.

[8] Targets should not be seen as substituting for the entire range of priorities of the health system, but should focus on what is currently deemed most important.

[9] Given the dynamics of health systems, it would seem prudent to set hard-edged targets in areas where professionals, managers and the public broadly agree that there are problems that need attention.

6 Targets should have a single, specific objective that is reasonably (not necessarily entirely) within the power of the relevant institution to influence. It should be clear to which agency or agencies the target relates. Where a performance target relates to more than one agency, joint targets can be set as long as it is clear which agency has 'lead' responsibility (and possibly control of the overall budget for meeting the target).

7 Target selection and the relative weight given to different targets in performance management is *always* a political process and requires political leadership. In a tax-financed system, it is entirely appropriate that central government ultimately determines the pattern of targets since it is accountable for the resources deployed on behalf of the public.

8 Targets do not have to be quantitatively weighted and aggregated into a single measure of performance (e.g. a 'star rating' or index of efficiency) to be useful for performance management[10]. Compound measures of performance are more likely to be valid and helpful in relation to public utilities rather than more complex fields such as health care (Smith and Street, 2005) though they do focus public and media attention on performance issues and sharpen underlying incentives. Index measures can be misleading and encourage 'gaming'.

9 Targets should be related to sanctions and rewards, but these must be proportionate to the service improvement or lack of improvement, and sensitively implemented (e.g. it makes no sense to sanction a manager for narrowly missing a performance target if substantial service improvement has nonetheless been demonstrated). Sanctions and rewards should not be used to signal the relative importance of individual targets since this encourages perverse behaviour. No one should be dismissed for missing a single target given that targets can only relate to specific aspects of performance, by definition, and there are likely to be 'grey' areas where performance is not entirely under the control of a specific organisation. Instead, sanctions and rewards should relate to overall performance in target areas even if performance is not measured through a single index since there will be inevitable uncertainties associated with the values recorded on individual measures.

10 Targets must be an integral part of any performance management, audit or inspection regime and should not be confused with other approaches to performance improvement such as peer benchmarking.

11 In tax financed systems, targets should be national except where it is evident that specific institutions or geographic areas have particular problems where it may be appropriate to agree local targets drawn from a domain of national priority areas or to quantify national targets differently (e.g. lower the percentage improvement required within a specific time for those institutions which are likely to face the greatest challenge in improving performance) (Bevan, 2006). Another approach is to set general targets at national level with successively more detailed targets below national level. The risk with this is that the national targets could become vague, thereby weakening accountability.

[10] The decision in the English NHS to produce an overall 'score' for each provider and purchaser organisation was designed to focus senior management and governance board attention of performance issues since it allowed the centre to publish 'league tables' of individual indicators and overall 'scores' which were picked up by the media. An effective target regime can be established without generating simple, single number scores for institutions or services. Index measures suffer from the difficulty of deciding which indicators to include and what weight to give to each.

12 Targets backed up by incentives will generate some undesirable behaviours and effects. These should be expected rather than seen as a fundamental threat to the policy and plans should be put in place as soon as possible to mitigate or eliminate them (e.g. by altering the way target performance is measured). Bevan and Hood (2006) suggest two other ways to reduce the odds of significant 'gaming': independent auditing and reporting of performance data; and the introduction of some uncertainty into the assessment of performance to reduce the odds of certain kinds of managerial 'gaming' (e.g. managers temporarily putting extra staff to work in a particular area when it is known that performance data are to be collected). They raise the intriguing possibility of varying the range of targets and other performance indicators used to assess an institution's performance each year (Bevan and Hood, 2004). They give as an analogy the relationship between the syllabus of a course and the topics from the syllabus covered in the related unseen examination, designed to reduce strategic behaviour by students.

8 Implications of the English NHS experience for New Zealand's public health system

The system of 'targets' (i.e. performance goals backed up by a strong incentive regime) in the English NHS was one strand in what Leatherman and Sutherland (2003) have described as, 'The most ambitious, comprehensive and intentionally funded national initiative to improve health care quality in the world.' This initiative embraced a wide range of different approaches to improving performance in line with what is known about improvement in health care organisations. As Ham (2003, p1978) puts it:

> 'Accepting that influences on clinical practice are many and varied, no one approach to improvement of performance is likely to be sufficient. Rather, several interventions are needed, including educational initiatives, use of opinion leaders, peer-review mechanisms, and financial and other incentives.'

While New Zealand uses many of the same approaches, there has been no comparable use of targets linked to incentives in New Zealand despite the similarities between the two public health systems in terms of central government funding, strategic objectives and organisational accountability. Though the New Zealand Health Strategy and related Strategies set a large number of objectives these tend to be aspirational, and are usually not quantified or time-limited. Thus the objectives are set out in the Strategy in terms of:

- reducing smoking
- improving nutrition
- reducing obesity
- increasing the level of physical activity
- reducing the rate of suicides and suicide attempts
- minimising harm caused by alcohol and illicit and other drug use to individuals and the community
- reducing the incidence and impact of cancer

- reducing the incidence and impact of cardiovascular disease
- reducing the incidence and impact of diabetes
- improving oral health
- reducing violence in interpersonal relationships, families, schools and communities
- improving the health status of people with severe mental illness
- ensuring access to appropriate child health care services including well child and family health care and immunisation.

However, these objectives are not formal 'targets' backed up by performance-related consequences. What are described as 'targets' exist in some areas, but these are not linked to managerial incentives so are not strictly 'targets' in the sense described here. Performance 'league tables' are not published.

The evidence presented above of the positive effect of 'targets' in the English NHS suggests that there is at least a case for considering some use of 'targets' in the New Zealand public health system in a small number of areas where there is a reasonable consensus that current performance can and should be improved, and that appropriate targets can be set and validly measured. This is not to say that the system should rely exclusively on such 'targets' to generate improvements in performance. 'Targets' should be seen as one among a range of approaches.

However, there are a number of issues in relation to the sparing use of 'targets' which would have to be taken into account in designing an effective system of targets and incentives. Some of these issues relate to the differences between the New Zealand public health system and the English NHS, others are generic to the use of performance targets with associated incentives. On the generic side, in addition to possible adverse side-effects such as some distorted behaviour, ineffective responses and fraud, and the need to put in place countervailing measures if targets are linked to relatively high powered incentives, there is the risk that a 'target' regime may reduce the internal motivation of professional staff (Smith, 2005). The costs of heavy reliance on external motivation through 'targets' may be considerable although much depends on the degree to which the relevant professionals agree or can be persuaded of the importance of pursuing the particular targets (e.g. it has been argued that one reason why waiting times were relatively neglected as a policy issue in the NHS was that clinicians were much less affected by their consequences than patients and so less concerned to shorten them than they were motivated to improve other aspects of their services such as improving the quality of care for those treated, but that the government succeeded in convincing a significant proportion of the professionals that the future of the NHS depended on drastically improving its responsiveness).

This tends to suggest that national 'targets' should be used sparingly and that, as far as is consistent with government responsibility and accountability for the use of large amounts of taxpayers' resources, they should engage the interest and commitment of health professionals as well as being directed at managers (arguably, the English regime was mostly (perhaps excessively) directed at incentivising managers). In this regard, the English experience in the cancer field is instructive in that the impetus for service change came from setting improvement 'targets' (e.g. that no patient should wait more than two weeks to see a specialist when referred by their GP with suspected cancer) through a relatively 'top-down' problem identification process, but the changes in response

('modernisation') were developed and implemented through regional 'collaboratives' or networks of organisations and professionals involved in all aspects of cancer screening, diagnosis, treatment and care. This process was led from the centre by the appointment of a so called 'cancer tsar' (the NHS National Cancer Director), Prof Mike Richards, a medical oncologist. In this way, a balance was struck, at least in the cancer field, between 'top-down' external scrutiny and accountability (through tracking progress towards 'targets' in terms of post-treatment cancer mortality rates), and reliance on more 'bottom-up' professionally led change (through staff themselves reconfiguring services and referral processes).

In general, it is likely that local clinicians will have interests and priorities for their services which differ appreciably from central government's systemic targets (Chang 2006) unless efforts are made to convince the relevant clinical communities of interest of the importance of the targeted area in terms that they can relate to. This does not necessarily involve accepting professionals' definitions of problems and priorities (given the gains that can be made by the use of targets), but it does involve recognising that the lower the degree of 'fit' between government targets for the system and the internal goals of professional providers, the greater the likelihood of organisational resistance to externally imposed requirements (Oliver, 1991). Broadly, clinicians in the relevant areas have to believe that meeting the target will benefit their patients without imposing excessive costs on the clinical community.

In designing targets and incentives, there are also the questions of how to set the standard embodied in any target and what to reward/penalise. Should targets be set in terms of raising average performance across a group of institutions, or should they be based on an expected level of *improvement* in performance irrespective of the starting point, or should they relate to how far from a specified standard or the group average an institution starts (i.e. so that improvement from a low base might count for more than the same improvement from a higher level of performance)? Further, should targets apply to all institutions or only to those performing below a specified level? Answers to these questions depend, in part, on the overriding goal of the target regime – whether it is to raise the average level of performance irrespective of which institutions contribute, or whether it is to raise the average by raising performance at the bottom end of the distribution, or whether it is to narrow the variation in performance while improving the average.

Evidence indicates that targets should be set, or at least rewarded, generally in terms of *improvements* in performance rather than the attainment of a particular level, since the latter tends to signal to average and above-average performers that they have nothing to accomplish and only sends out strong incentives to poor performers. In addition, a focus on improvement does not disadvantage organisations (in this case, District Health Boards (DHBs) and Primary Health Organisations (PHOs)) which serve more deprived or higher need populations since they will be rewarded for improvement irrespective of where they start from. This approach also means that it is less critical than it would otherwise be to be able to adjust any performance measures exactly to take into account differences between the populations of DHBs or PHOs. However, there may be situations where targets embodying an absolute standard are preferable (e.g. where performance below a particular level is widely regarded as unacceptable and the focus is on improving the performance of those who fall below the standard irrespective of where they start from and irrespective of the characteristics of their population served).

In the context of a much smaller health system than the English NHS, with only 21 DHBs (though 81 PHOs), separating real variations in performance from apparent variations

attributable to small numbers of events will be important in order to maintain the credibility of the approach. Any variations in performance which attract sanctions and rewards have to be valid (e.g. calculated on the basis of moving averages over a number of years and/or presenting 95% confidence intervals around any point estimates).

Another important design issue is to set targets that are within the scope of the relevant agency or agencies to influence if not entirely remedy. For example, while a health problem may be a major contributor to the burden of disease, it may not be amenable to action on the part of the health sector or there may not be adequate knowledge of how the health system can best respond. It would be important, particularly at the beginning of the process not to set strongly incentivised targets which risk undermining confidence in the approach.

Other issues which have to be determined in developing a system of incentivised targets include what rewards/sanctions attach to good/poor performance (and/or little or no improvement in performance), whether or not there should be any local as against national targets, and whether or not progress should be measured by an organisation other than the organisation which sets the targets. Of these, the most sensitive is the choice of financial and non-financial rewards and sanctions facing DHBs and other health sector organisations. At present, little use is made of either financial or non-financial rewards and sanctions, yet it is well known that simply collating and monitoring performance information is unlikely to produce substantial performance improvement unless supported by incentives. The only financial reward available to DHBs currently is advance payment on a monthly basis from the Ministry of Health to those DHBs that are performing well financially. It is possible that a similar advance payment could be made available to DHBs that make non-financial performance improvements in target areas. However, this is a very limited reward, particularly given that it would offer nothing to those DHBs that are already performing well financially.

Another possibility in the New Zealand context, where equity of funding and access to services between DHBs are high priorities (as they are in England), might be to offer high performers a share in a staff development fund. However, the vertical integration of planning and funding with hospital provision in the DHB model makes it more difficult than in the English NHS to reward providers versus purchasers for their contribution to performance improvements since publicly owned providers are part of the DHB. There is some scope to use peer and public recognition as an incentive if performance trends are publicly reported. Again, the ability to do this is reduced in the New Zealand context by the fact that purchase and provision of hospital and public health services are the responsibility of the same organisation making it less likely that a DHB could or would use provider performance data to alter its pattern of purchasing. Similarly, many DHB hospitals (if not other service providers) are local monopolies further reducing (but not necessarily eliminating) the ability of the centre to use comparative performance assessment to encourage performance improvement.

There is also the question of whether there should be any external assessment of why performance of an institution or area is poor in order to be able to offer advice versus leaving this entirely to local initiative. In the English NHS, the NHS Modernisation Agency developed to provide a management consultancy service separate from the Department of Health to work alongside local providers to help them with problem identification, development of solutions and implementation in order to meet targets (among other performance issues identified at local level). The Agency also publicised case studies of improvement so that other NHS organisations could benefit from the achievements of

leaders in the field. It is possible that the Ministry of Health could contribute in this way since it is unlikely that a separate agency could be justified in the New Zealand context.

9 Conclusion

The results achieved in the English NHS through a regime of targets and related incentives suggests that rather than rejecting them because of their admitted drawbacks, there is scope to use targets and related incentives sparingly to improve performance in the New Zealand public health system. The challenge is to learn from the English experience and elsewhere in order to maximise their social benefits and minimise their costs as part of a range of different approaches to performance improvement.

10 References

Alvarez-Roseté A, Bevan G, Mays N, Dixon J. (2005) Effect of diverging policy across the NHS. *BMJ* 331: 946-50

Bevan G. (2006) *Setting targets for health care performance: lessons from a case study of the English NHS.* Paper presented at the 4th NIESR Public Service Performance Conference. London: British Academy, 20 January 2006

Bevan G, Hood C. (2004) Targets, inspections and transparency. *BMJ* 328: 598

Bevan G, Hood C. Have targets improved performance in the English NHS? *BMJ* 2006; 332: 419-22

Carter N, Klein R, Day P. (1995) *How organisations measure success: the use of performance indicators in government.* London: Routledge

Chang L. (2006) Managerial responses to externally imposed performance measurement in the NHS: an institutional theory perspective. *Financial Accountability and Management* 22: 63-85

Collins P, Sibson J, Shotas J, Smith P, Thornton S. (2005) *To the point: a blueprint for good targets.* Report of the Commission on Targets in the Public Services. London: Social Market Foundation

Department of Health. (2005) *Chief executive's report to the NHS.* London: Department of Health

Ham C. (2003) Improving the performance of health services: the role of clinical leadership. *The Lancet* 361: 1978-80

Hauck K, Street A. (2006) Do targets matter? A comparison of English and Welsh national health priorities. *Health Economics*, forthcoming http://www.york.ac.uk/inst/che/pdf/streetperformance.pdf

House of Commons, Public Administration Select Committee. (2003) *On target? Government by measurement.* London: The Stationery Office

Le Grand J. (2004) *Choice, voice and the reform of public services.* Annual LSE Health and Social Care Lecture. London: London School of Economics and Political Science

Leatherman S. (2005) *Improving performance in health care systems: facts, fallacies and forward thinking.* Presentation to Cambridge Health Network, 29 September 2005

Leatherman S, Sutherland K. (2003) *The quest for quality in the NHS: a mid-term evaluation of the 10-year quality agenda*. London: The Nuffield Trust

Jacobs R, Martin S, Goddard M, Gravelle H, Smith P. (2006) Exploring the determinants of NHS performance ratings: lessons for performance assessment systems. *Journal of Health Services Research & Policy* 11: in press

Joint Commission on Accreditation of Health Care Organisations. (1990) *Primer on indicator development and application.* Chicago: JCAHO

Mannion R, Goddard M. (2001) Impact of published clinical outcomes data: case study in NHS hospital trusts. *BMJ* 323: 260-3

Mannion R, Goddard M. (2002) Performance measurement and improvement in health care. *Applied Health Economics and Health Policy* 1: 13-23

Oliver C. (1991) Strategic responses to institutional processes. *Academy of Management Review* 16: 145-79

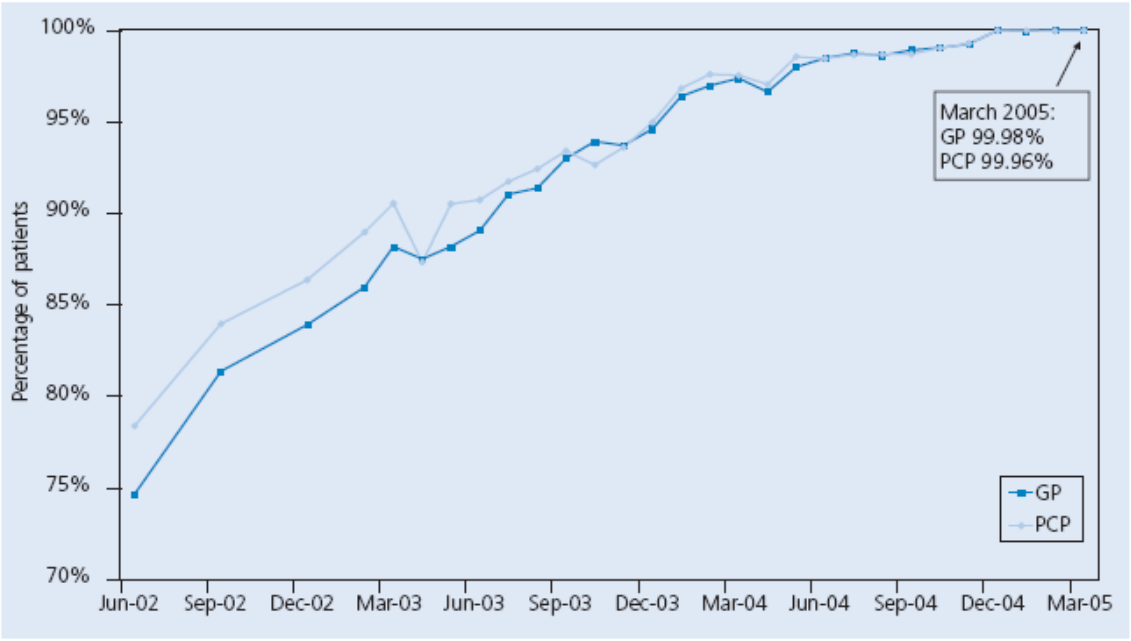
Schyyve P. (1995) Models for relating performance measurement and accreditation. *International Journal of Health Planning and Management* 10: 231-41

Smith PC. (2005) Performance measurement in health care: history, challenges and prospects. *Public Money & Management* 25: 213-20

Smith PC, Street A. (2005) Measuring the efficiency of public services: the limits of analysis. *Journal of the Royal Statistical Society A* 168: 401-17

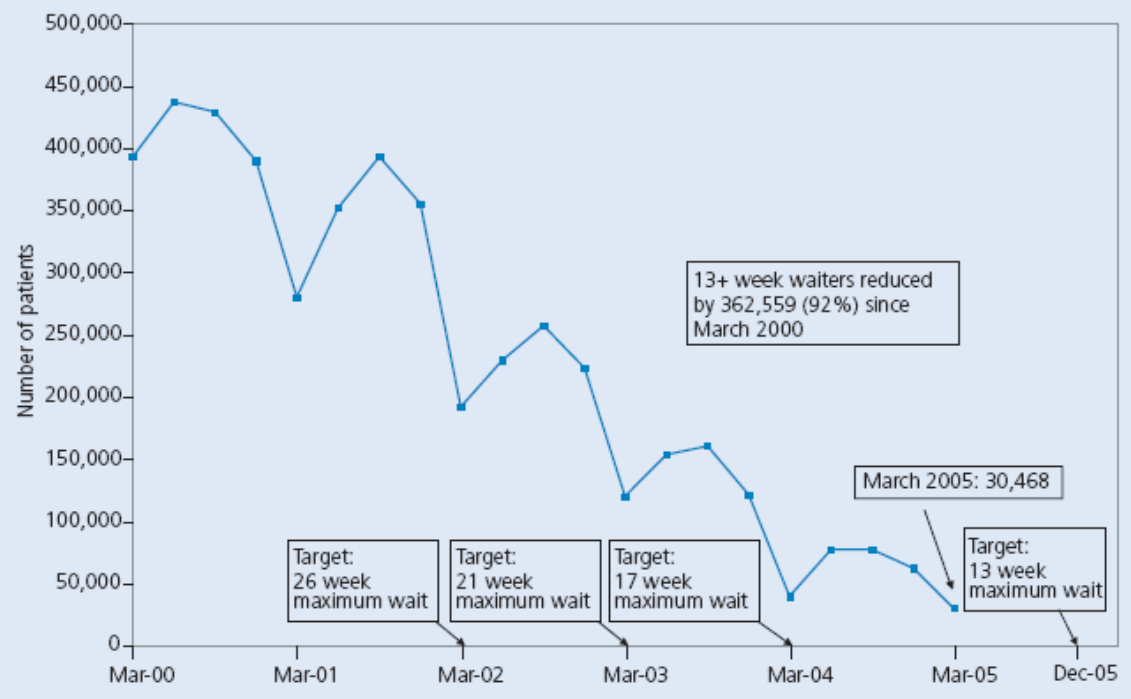
Appendix 1: Further trends in waiting and waiting lists in the English NHS

Figure 8: Percentage of patients with access to a GP within 48 hours or to a primary care professional (PCP) within 24 hours



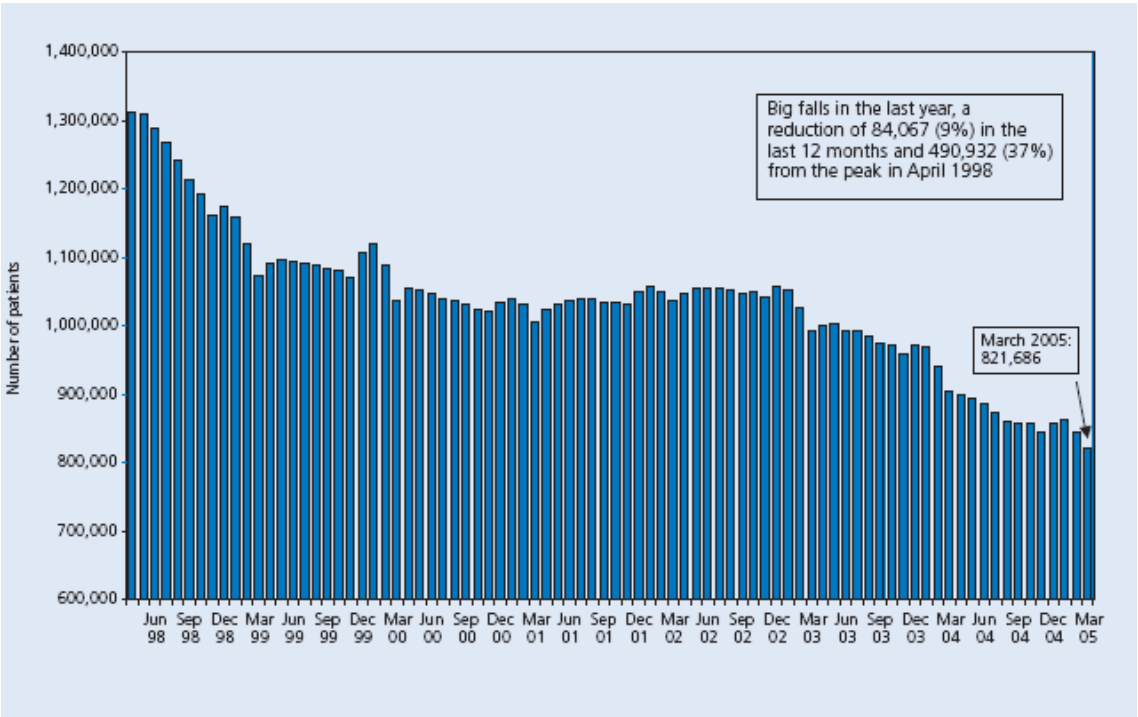
Source: Department of Health (2005) *Chief Executive's Report to the NHS*. London: Department of Health

Figure 9: Patients waiting over 13 weeks for an outpatient appointment



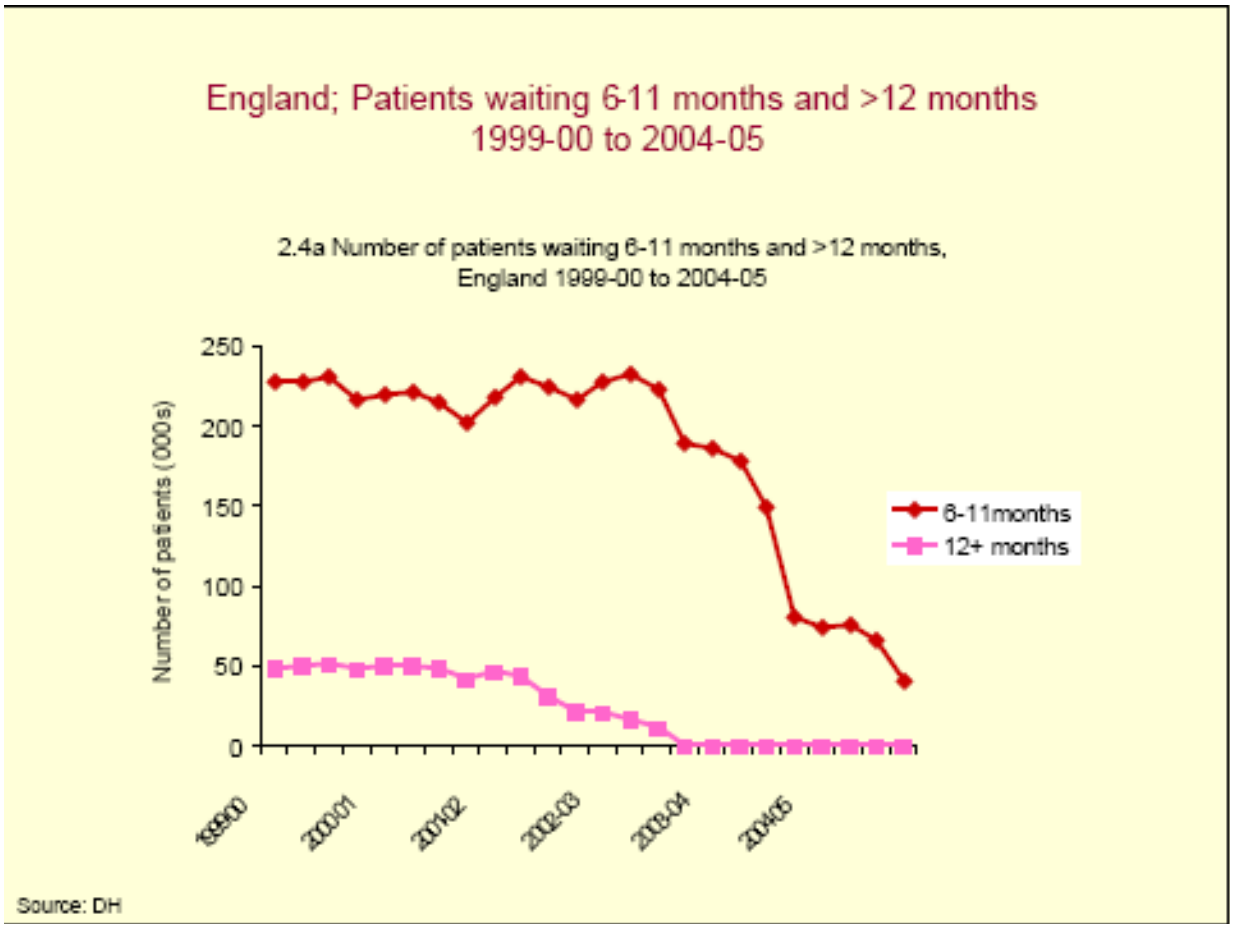
Source: Department of Health (2005) *Chief Executive's Report to the NHS*. London: Department of Health

Figure 10: Inpatient waiting list numbers



Source: Department of Health (2005) *Chief Executive's Report to the NHS*. London: Department of Health

Figure 11: Inpatient waiting times (months waited)



Source: Leatherman (2005)