

Gaming and Targets in the English NHS

Alex Mears

Faculty of Health and Social Care, London South Bank University

*Corresponding Author: alex.mears@nhs.net

Copyright © 2014 Horizon Research Publishing All rights reserved.

Abstract Process-driven performance targets were used in the English NHS in the late 1990s and early 2000's to drive improvement. There have been clear successes, and well-documented instances of perverse incentives leading to negative behaviours, such as data manipulation and resource diversion. Recently, policy in England has moved away from targets, focusing more on locally-negotiated metrics. The literature was searched for relevant materials, from a broad social science perspective. Probative work from all disciplines was brought together to give an overview of targets, perverse incentives and the routes to minimising such unintended consequences, and to enable conclusions to be drawn. The consequences of targets are well-documented, but less well understood. Routes are available to mitigate the potential for perverse incentives, within the design and interpretation phases and also in terms of sanctions for gamers. These have largely not been followed, rather the political fashion has moved away from targets altogether. It is hypothesised that this will lead to an increase in waiting times, although this has yet to fully manifest. While targets may swing in and out of fashion, the psychological drivers remain: wherever a measure is imposed that carries with it a consequence, there will be a perverse incentive. Process-driven targets in healthcare have supporters and critics. Their heyday in the English National Health Service was primarily during the New Labour Government of the late 1990s and early 2000s, with the combination of process-heavy targets and high levels of incentive leading to not only high levels of compliance and improvement, but also well-documented instances of gaming behaviour. Proponents of the use of targets point to notable successes in the reduction of waiting times where key targets were introduced (such as ambulance Category A (life-threatening cases) response times). Opponents raise a number of issues around the negative behaviours that targets give rise to, for example the diversion of resource to the measured area and the distortion of clinical priorities. Recently, it would appear that health policy has moved away from targets as a measurement system; in this paper we argue that if it is hoped that this will prevent perverse incentives and gaming arising, then this is unlikely to succeed: wherever performance or quality are measured, then there will be an incentive to manipulate the data or circumstances to give the appearance of compliance. We conclude with a consideration of the strategies available to reduce or mitigate

gaming through how measures are designed, interpreted and how gaming is perceived.

Keywords Gaming, Perverse Incentives, Performance Measurement, Targets

1. Background

In the years following the election of the Labour government in 1997, targets and star ratings for NHS organisations combined an effective system of measurement with high-powered incentives; two outcomes are well-documented: a substantial decrease in waiting times in the intervention areas targeted, and the burgeoning of 'gaming' behavior.

The evidence suggests that targets can be very effective at incentivizing desired behavior, and achieving desired outcomes, which can be a benefit for the health system (for example through shorter waiting times, or improved outcomes); yet targets can also give rise to undesirable behaviors and therefore the benefits are not realized, or in fact outcomes are worse. The problem, then, is to design a target-based measurement system that avoids the undesirable behavior while retaining the benefits.

The most obvious form of undesired behavior is 'gaming'. This comes in many forms, but the linking theme is a manipulation of the situation or the data to make performance appear better than it actually is. An often-cited example is the removal of wheels from a trolley and redesignation as a 'bed' to avoid breach of an emergency room waiting time target (British Medical Association, 2000). Gaming is most prevalent where a performance measure is attached to tangible consequences, either negative or positive such as reputational damage or enhancement (Bevan & Hamblin, 2009). In England, the star rating system was known for inflicting reputational damage on those organizations scoring poorly (Department of Health, 2001). In the literature on the effects of star ratings one of the acknowledged negative effects was gaming behavior (Mannion et al, 2005).

This paper uses a modified case study methodology to explore the concepts of performance measurement, targets, gaming and incentivisation. The experience of the English

NHS is unpacked and used to illustrate the concepts introduced, and to further consider how this experience could be used to design a more effective target-based measurement regime. The pitfalls of moving away from a target-based system are also considered: the implications of this shift, in terms of the perverse incentives created and whether gaming behaviour will become less of a concern or merely change form and focus. Finally, a consideration is made of how to make best use of the positive aspects of targets whilst avoiding the negative in terms of gaming.

2. The Rise of Targets

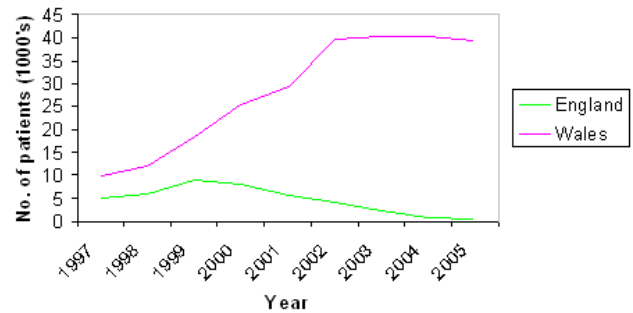
In recent years, waiting times for NHS treatment had historically been high and much criticised because of this. The length of time that patients had to wait for treatment became a cause of major concern to the media and public in the 1980's and 1990's, leading in 1991 to the introduction by the Major government of the Citizen's Charter (Prime Minister, 1991), which for the first time included a maximum waiting time for certain kinds of treatment (although diagnostic waits were excluded).

In 1997, the New Labour government of Tony Blair made the reduction of waiting times a keystone of policy (Labour Party, 1997), leading to the publication of the NHS Plan in 2000 (Department of Health, 2000), which introduced more waiting time targets and additionally public service agreements (PSAs) that added further and shorter waiting times, and targets in health and social care (eg health inequalities, smoking cessation etc). Drawing these indicators together, the introduction in 2001 of aggregated star ratings augmented the importance of the individual targets that now contributed to an overall score for the organisation. While other indicators were included within the calculations that produced the star rating, the key drivers were the waiting times.

Targets have without doubt led to some quantifiable

improvement successes in those areas within scope, notably in waiting time decreases (Appleby & Harrison, 2005). A recent report shows the stark differences between improved performance in England (where targets existed and good performance was incentivised) and the other nations in the UK (where perverse incentives existed- poor performing organisations received extra funding; Besley et al, 2008). The devolution of health policy from the UK government seat in Westminster to the Senedd (Wales), Stormont (Northern Ireland) and Holyrood (Scotland) bodies facilitated a natural experiment. While all regimes used targets to measure performance, only England introduced incentives for good performance and censure for poor performers; other nations retained the perverse incentive of being perceived to reward poor performers (Connolly et al, 2010). The divergence between the incentive-based regime and others is stark, and can be seen by looking at comparative data for some key indicators.

Data from GP referral to outpatient waiting times shows a high differential between England and Wales (figure 1).



In another targeted area, the incidence of MRSA and C.difficile in acute hospitals in England has dropped significantly within the period that they have been subject to a high-profile target (HPA, 2010; figures 2 & 3).

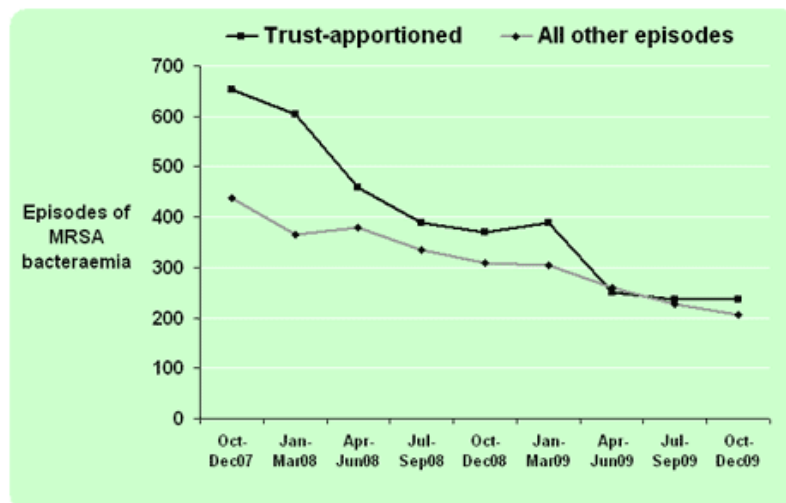


Figure 2. Episodes of MRSA Bacteraemia, 2007 to 2009

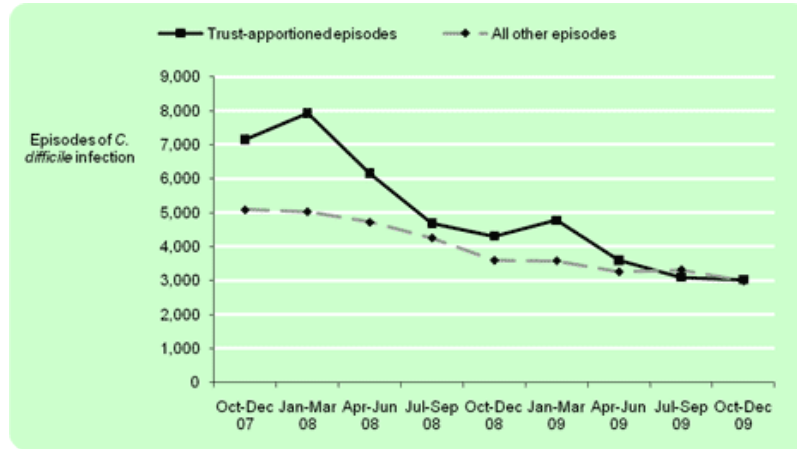


Figure 3. Episodes of *C. difficile* infection, 2007 to 2009

Finally, it is worth considering the impact of the introduction of incentivised targets in ambulance response times, as highlighted in Connolly et al (2010) – see figure 4).

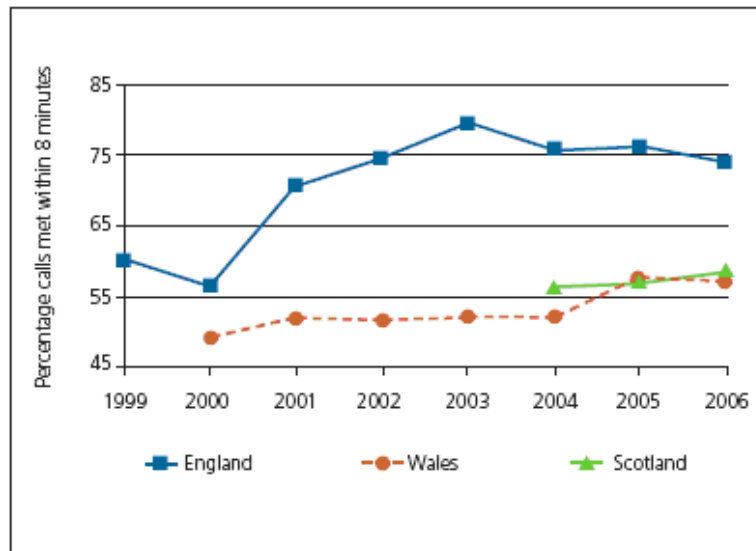


Figure 4. Percentage of Ambulance Calls Answered within 8 minutes in England, Wales and Scotland.

Response times in England improved enormously following the introduction of incentive-based targets. This was achieved through a panoply of sound improvement planning and unfortunately gaming through direct data manipulation (Bevan & Hamblin, 2009). It should be noted that one element of gaming uncovered here involves improving an already good score (‘gilding the lily’; Bevan & Hamblin, 2009).

Evidence suggests, then, that targets do work, that they lead to improvement in the areas subject to them. But this approach has lost popularity recently, despite these successes.

3. The ‘Bonfire of the Targets’ is Lit

Targets have become unpopular for a number of reasons, criticism coming from many quarters: clinicians, politicians,

NHS managers, and press. Some of the censure is politically motivated, some from genuine concern. Some is based on unsubstantiated prejudice and just wrong; there are some valid, acknowledged disadvantages to a targets based regime, however.

1. Resource diversion (the ‘synecdoche’, Bevan & Hood, 2006)

The introduction of a target in one area of activity will naturally encourage the diversion of resource to that area, at the expense of the unmeasured. This effect is more profound where incentives are attached to the target, as was the case in English health policy as described above. An extreme example of this (albeit with a very different kind of target to those discussed above) can be seen in events at Mid Staffordshire NHS Foundation Trust where pursuit of financial targets led to patient neglect (Francis Enquiry Report, 2010).

2. Distortion of clinical priorities

Targets are not sensitive to the relative importance of different conditions, unless specifically addressed within the construction. Thus, the waiting time for an ingrowing toenail has as much importance and a hip replacement to the target.

3. Targets encourage the organisation to ‘face the wrong way’

Seeking to satisfy the objective target becomes paramount to the organisation, with the patients or clients it serves being of secondary importance only (Chapman, 2007).

4. Productive time is lost

The introduction targets will inevitably draw resource away from frontline activity (through the necessity to collect, collate and interpret data; Chapman, 2007).

5. ‘Hitting the target and missing the point’ (Bevan and Hood, 2006)

Diverting patients into a ‘pre-assessment unit’ prior to starting the A&E 4 hour waiting time helps achieve the target but does not address the issue it was introduced to combat, long patient waits.

6. Targets can lead to unintended consequences (such as gaming)

Gaming is the deliberate distortion or manipulation of information or circumstances to make performance or quality appear better than it is. A target will encourage organisations to look for routes to giving the appearance of high performance, in order to reap the associated rewards. Gaming can be advantageously coding patients to maximise income to direct falsification of records to achieve a waiting time, with many other examples documented.

Professor Lord Ara Darzi’s ‘Next Stage Review’ of the NHS (Department of Health, 2008a) emphasised a shift further away from targets and centralisation towards local performance measurement. Following the 2010 general election the Secretary of State for Health, Andrew Lansley, is responding to Conservative calls for a ‘bonfire of the targets’ (Guardian, 2010).

In the months following the 2010 election, we have seen the abolition or amendment of what might be called three of the New Labour flagship targets: 48 hours for a GP appointment, 18 weeks from referral to treatment target, and reduction of the 4 hour A&E waiting time threshold from 98% to 95%; these changes are contained within the DH operating framework for 2010/11 (DH, 2008b).

Hand in hand with the move away from centrist target-based measurement has been a shift away from process-based measures towards primarily clinical outcomes, as contained within the new Outcomes Framework (DH, 2010). One final change is a more punitive approach to unnecessary readmissions. Under the Operating Framework, and readmission within 30 days of discharge will not be paid for, effectively a financial penalty. This surely can be categorised as another kind of centralised target?

These changes represent a sea change in measurement policy, effectively ending target-based assessment.

4. A Redesign of Targets to Meet the Criticism: Getting Rid of the Bathwater, Keeping the Baby

It could be argued that in any effective system of measurement will result in perverse incentives (Merton, 1936). In fact, the existence of gaming behaviour might be seen as an indication that a measurement system is being taken seriously (change of status quo manifested through behaviour). Organisations will only engage in gaming behaviour where the incentive is powerful and the measure effective in discerning good from poor performance. Crucially, in addition, there must be a perception that any gaming behaviour will not be detected or if it is, it will be ignored.

It is our contention that gaming is a reaction to effective measurement, and the removal of targets will therefore not eradicate it. Gaming will persist, in a different form, in response to any centrally mandated measurement upon which incentives are based, be they direct such as pecuniary advantage or disadvantage, or indirect in the form of lost kudos.

There is a sense that the new coalition regime, in its enthusiasm to rid healthcare of targets has thrown out baby and bathwater. We contend that through an understanding of the faults of target-based measurement, an effective route to incentivisation can be achieved without the damaging and distorting gaming behaviour.

In such a complex system as healthcare with many and varied complex measurement systems at work, removing gaming completely would be virtually impossible, and involve the invocation of a draconian system of measurement. To minimise it, however, is possible, through three routes: measure design, measure interpretation and behaviour change.

5. Measure Design

Bevan and Hood (2006) hint at the importance of ‘designing out’ potential gaming in how measures are formed. The most important of these is the avoidance of large incentives coupled with single measures. The more complexity within a measurement system, the less likelihood that gaming will take place, simply because it becomes too burdensome to do so.

In considering the effect of targets for Ambulance response times, Bevan and Hamblin (2009) highlight the importance of using a balance of measures from the Donabedian framework (structure, process, and outcome). By centring on just processes, as happened with ambulance response times, Goodhart’s second law is invoked: the measured entity changes its behaviour to meet the target, thus disconnecting the desired behaviour and the target linked to it (Goodhart, 1984). The overriding driver becomes

the target (ambulances responding within 8 minutes), not the reason it was introduced (to get ambulances to patients sooner and save lives). In addition, the authors further highlight the importance of not assessing organisations in isolation; measures should operate together to avoid gaming by trust A having an adverse effect on trust B (for example an acute trust making an ambulance wait until it was confident that an admitted patient would meet the 4-hour A&E waiting time; Bevan and Hamblin, 2009).

An as yet unexplored route to avoidance of gaming is to employ the principles of mechanism design (Binmore, 2007). This part of game theory suggests that where the activities of organisations (ie trusts) are not within the knowledge of their managing body (ie DH), then there is value in adjusting central policy such that the goals of the trusts are inclined to be set in such a way that their stable position (equilibrium) aligns with the goals of the DH. This effectively means that measurement has to be carried out in such a way that there is more incentive to improve performance than to game- the incentive to game is gone, as it is economically more viable to improve performance. This would prove far too difficult for an organism as complex as the NHS, but the underlying principles apply. This broadly fits in contextually with a finding from Fisher and Downes (2008) that gaming is far less likely where the assessed organisations do not feel that targets are measured primarily for the good of governmental reputation. The previous government recognised this through the de-targetisation proposed in the Next Stage Review (Department of Health, 2008a).

6. Measure Interpretation

The second set of paths to minimising gaming is through how performance measures are interpreted and used in regulation and performance management. A number of possible methods exist:

- Non-disclosure

One simple route to avoid some forms of gaming is to not disclose details of the measurement process to the assessed organisation (in the language of Bevan and Hood, 2006, “introducing some randomness into monitoring and evaluation in order to limit gaming”). Incomplete information makes the gaming process considerably more difficult and drains more resource, meaning that many organisations would be far more reluctant to engage.

- Moderating indicators

A moderating indicator can be used almost as a detection mechanism for gaming behaviour, and would act to disincentivise it. A hypothetical example raised above is where an acute provider is meeting its 4-hour A&E wait by making ambulances wait until the staff were confident that the patient within could be seen in 4 hours. A measure of ambulance turnaround could be used to moderate the 4-hour wait. An acute trust meeting its 4-hour target, but with consistently high ambulance turnaround figures could be

moderated (Bevan and Hamblin, 2009).

- Microframeworks

A microframework works in a similar way to a moderating indicator, instead using a framework built from a small number of indicators to detect gaming. An example would be to use staffing ratios in A&E before and during monitoring for the 4-hour A&E wait to moderate that wait target (ie to detect if far more staff were present during the monitoring period).

- Structured frameworks

As stated in Bevan and Hood (2006), single measures attract high levels of gaming (up to 33%). The more complex the measurement system, the harder and more pointless it is to game. Therefore, using a structured framework to assess, with levels of complexity, will minimise gaming.

- Wider measurement

Measuring more aspects of performance would tend to reduce resource diversion behaviour, although would add to the measurement burden.

7. Behaviour Change

A key route to reducing the appetite for gaming is to ensure that behavioural drivers are coalescing away from this course of action. There are several routes through which this may be achieved.

- Sanctions for gamers

Fisher and Downes (2008) make the point that in many cases, the penalty for being discovered engaging behaviour is not serious. This is supported by the findings of the NAO (2001): even senior executives caught gaming can generally remain working in the NHS, or are allowed to slip away with dignity. Low-level gaming is widely tolerated and in some organisations encouraged as part of the culture (Jensen, 2003). What might help to reduce this behaviour is to introduce some more serious sanctions for those found engaging in gaming behaviour, at organisational level and possibly with personal legal effect (ie prosecution). Gaming could trigger regulatory activity (an audit of data handling for example).

- Culture of unacceptability & peer pressure

An important route to reducing gaming is to engender a culture of unacceptability. At the moment, some organisations have embraced a culture of gaming, for a number of reasons (Jensen, 2003). What might be instructive is to require all chief executives to introduce a ‘zero-tolerance’ policy against gaming, in order to make such behaviour become a taboo. Peer pressure would be a useful mechanism to prevent low-level gaming. It might be hypothesised in addition that the current climate may well shift to mean that gaming is not seen as acceptable any more.

8. Gaming in the New World

In this last section of the paper, we examine some of the

remaining and new indicators and hypothesise on how these might be gamed.

The Outcomes Framework

The Outcomes Framework published to support the DH white paper on NHS reform outlines proposals for quality measurement in the NHS (DH, 2010). The white paper does not include a great deal of detail as to the nature and substance of the framework that sits beneath the highest level measures. It is clear that there is no room for targets as we have known them, with an emphasis on outcome measures. While this should render gaming more challenging, what we do know is that where a measure matters, there is an incentive to game it. We can predict that gaming will not be eliminated by this new measurement regime, but merely change form.

Mortality outcome measures

The Hospital Standardised Mortality Ratio (HSMR; Jarman et al, 2005) and its replacement the Summary Hospital-level Mortality Indicator (SHMI; NHS Information Centre, 2011) is an outcome measure that is therefore likely to remain in use. These Mortality Ratios (MRs) are used by a number of organisations (primarily Doctor Foster Intelligence and CQC, Aylin et al, 2009; Healthcare Commission, 2009) as a measure of performance or risk. The MR is the ratio of observed to expected deaths in an organisation, multiplied by 100. Thus, any reduction in observed deaths or increase in expected deaths will render the outcome more favourable to the organisation. There are a number of routes to manipulate either the observed deaths down or expected up to improve the SMR. For example, moving moribund patients from ward to another location prior to death, recategorisation of patients into HRGs not used to calculate the SMR. It is possible, however, that SMRs will fall out of favour as a measure of quality, after Sir Bruce Keogh's statement in his recent report (Keogh, 2013) said "However tempting it may be, it is clinically meaningless and academically reckless to use such statistical measures to quantify actual numbers of avoidable deaths"

30 day readmissions

The current Secretary of State has announced a new initiative that will penalise acute providers where a discharged patient is readmitted as an emergency case within 30 days with a 'related problem' (BBC website, June 2010).

This initiative is designed to ensure that patients are discharged in a timely fashion, with appropriate care plans in place- where this doesn't happen, readmission is considered to be more likely- readmission rates have increased considerably in recent years.

Such an initiative does, however, create the opportunity for perverse incentives to lead to new forms of gaming behaviour. It might be hypothesised that hospitals will have an incentive to hang onto patients longer than is clinically necessary; there is also an incentive to avoid admitting more seriously ill patients at all.

Payment by results

Payment by Results (PbR) is an activity based finance model for acute care providers based on their activity levels and case-mix. Patient admissions are categorised into healthcare resource groups (HRG) and providers are paid a national tariff for each HRG. The tariff for each procedure is based on the average cost of all NHS providers. The introduction of PbR shifted the commissioning of acute services away from block contracts, which were locally negotiated between providers and commissioners, to a national standardised system. Block contracts, in general, were historically based and adjusted up each year; such contracts included a lot of 'slack' and did not promote the efficient allocation of resources. Unlike PbR, prior contract agreements were sensitive to cost and manipulation by providers.

PbR has introduced an array of incentives to the acute care sector. For instance, there is an incentive to ensure that all patients are coded in the most advantageous way (that which pays most), regardless of the reality of clinical need (Rogers et al, 2005). Similarly, multiple stays will yield income for each additional one. Finally, there is a market incentive for providers to centre efforts and resource towards those interventions with the lowest cost/ highest return. It might be contended that this paves the way for plurality through ultra-specialisation. The key question is how these incentives can and will be managed to alter behaviour in a positive manner and avoid perverse behaviours. A watchful eye must be kept on the 'unintended consequences' of this payment mechanism.

9. Conclusions

Gaming and data manipulation is an unwanted, unintended consequence of performance measurement in healthcare. The target-based measurement system certainly created perverse incentives including gaming, but this is not a uniquely target-based phenomenon. Certain drivers are likely to reduce familiar event forms, but it is hypothesised that new forms will evolve. Through an understanding of what drives gaming at both the individual and organisational level, we can begin to formulate methods to prevent or at least reduce it.

Acknowledgements

No sponsorship, funding or affiliations.

The author would like to acknowledge the help and support received from:

Richard Hamblin, Director of Quality Evaluation, Health Quality and Safety Commission, New Zealand

Professor Gwyn Bevan, Professor of Management Science, London School of Economics and Political Science

REFERENCES

- [1] Aylin P, Bottle A, Hua Jen M, Middleton S. HSMR mortality indicators. London: Doctor Foster Intelligence, 2009.
- [2] BBC news website. Hospitals to face financial penalties for readmissions. See <http://www.bbc.co.uk/news/10262344> (last accessed 15th December 2010).
- [3] Besley T, Bevan G, Burchadi K. Accountability and incentives: The impacts of different regimes on hospital waiting times in England and Wales. London: London School of Economics, 2008.
- [4] Bevan G, Hood C. What's measured is what matters: Targets and Gaming in the English Public Health Care System. *Public Administration* 2007;84(3): 517-538.
- [5] Bevan G, Hamblin R. (2009). 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 (A)* 2009;172(1):1-30.
- [6] Binmore, K. Rules of the Game. *Prospect magazine* 2007 140, Nov 2007.
- [7] British Medical Association. Survey of Accident and Emergency Waiting Times. London: BMA, 2000
- [8] Chapman J. After Blair: Politics after the new labour decade. London: Lawrence and Wishart, 2007.
- [9] Connolly S, Bevan G, Mays N. Funding and Performance of Healthcare Systems in the Four Countries of the UK before and after Devolution. London: the Nuffield Trust, 2010.
- [10] Department of Health. The NHS Plan: a plan for investment, a plan for reform (2000). London: Department of Health, 2000
- [11] Department of Health. High quality care for all: NHS Next Stage Review final report. London: Department of Health, 2008
- [12] Department of Health. Operational plans 2008/09- 2009/10 (2008). London: Department of Health, 2008
- [13] Department of Health. Transparency in outcomes - a framework for the NHS. London: Department of Health, 2010.
- [14] Fisher, C., Downes, B. Performance management and metric manipulation in the public sector: a research report. *Business Ethics: a European review*. 2008; 17(3) 245-258.
- [15] Goodhart C. Monetary theory and practice. London: Macmillan, 1984.
- [16] Guardian. NHS waiting-time targets scrapped by Andrew Lansley. *Guardian* 21 June 2010.
- [17] Harrison A, Appleby J. The War on Waiting for Hospital Treatment: What has Labour achieved and what challenges remain? London: King's Fund, 2005
- [18] Healthcare Commission. Following up mortality 'outliers'. London: Healthcare Commission, 2009
- [19] Health Protection Agency. MRSA and C. difficile infections fall by a third: HPA press release. London: Health Protection Agency, 2010.
- [20] Jarman B, Bottle A, Aylin P, Browne M. Monitoring changes in hospital standardised mortality ratios. *British Medical Journal* 2005; 300: 329
- [21] Jensen M. Paying people to lie: the truth about the budgeting process. *European Financial Management* 2003; 9(3): 379-406.
- [22] Keogh, B. Review into the quality of care and treatment provided by 14 hospital trusts in England: overview report. London: Department of Health, 2013.
- [23] Mannion R, Davies H, Marshall M. Cultures for Performance in Health Care. Maidenhead: McGraw Hill, 2005
- [24] Merton, R. The unanticipated consequences of purposive social action. *American Sociological Review* 1936; 1(6): 894-904
- [25] National Audit Office. Inappropriate adjustments to NHS waiting lists Report by the Comptroller and Auditor General HC 452 Session 2001-1001: 19 December 2001. London: National Audit Office, 2001
- [26] NHS Information Centre. Indicator Specification: Summary Hospital-level Mortality Indicator. 2011. London: NHS Information Centre for Health and Social Care.
- [27] Labour Party. New Labour- Because Britain Deserves Better, Labour Party Manifesto 1997. London: Labour Party, 1997.
- [28] Prime Minister. Citizen's Charter, Raising the standard. London: HMSO, 1991.
- [29] Robert Francis Inquiry report into Mid-Staffordshire NHS Foundation Trust. London: HMSO, 2010
- [30] Rogers R, Williams S, Jarman B, Aylin P. "HRG drift" and payment by results. *BMJ* 2005; 330: 563.
- [31] Secretary of State for health. NHS Performance Ratings Acute Organizations 2000/2001. London: Department of Health; 2001