

THE UNIVERSITY *of York*

CENTRE FOR HEALTH ECONOMICS
YORK HEALTH ECONOMICS CONSORTIUM
NHS CENTRE FOR REVIEWS & DISSEMINATION

THE YORK SERIES ON THE NHS WHITE PAPER -
A RESEARCH AGENDA

**The NHS Performance Framework:
Taking Account of Economic Behaviour**

Maria Goddard
Russell Mannion
Peter C Smith

DISCUSSION PAPER 158

**THE NHS PERFORMANCE FRAMEWORK:
TAKING ACCOUNT OF ECONOMIC BEHAVIOUR**

by

**Maria Goddard
Russell Mannion
Peter C Smith**

August 1998

CENTRE FOR HEALTH ECONOMICS DISCUSSION PAPER SERIES

The Discussion Paper (DP) Series was established in 1984 to make research available quickly to a broad readership drawn from the academic community, professional health care staff and policy makers, both within the UK and internationally. Material published in DP form ranges from work in progress, to edited versions of completed project reports, from papers intended ultimately for journal publication, to those which seek to stimulate immediate debate. Hence the variation in length and complexity of DPs.

In the past, DPs have been subject to a minimum of editorial revision, it being principally seen as the responsibility of individual authors to ensure that appropriate standards were met. Since June 1996, a more formal system of internal peer-review has been introduced to consolidate the high levels of quality that have been achieved in the past. That system involves the refereeing of DPs in draft form, with any recommended revisions being verified independently by a senior member of staff.

DPs are not a substitute for publication in peer-reviewed journals. It is to be hoped that journal editors will recognise the nature of the DP series, and accept that this does not compromise the status of those papers published by the Centre which may later be submitted for consideration in more formal publications. Comments and suggestions on DPs encouraged and should be sent directly to the authors.

THE AUTHORS

Maria Goddard is a Senior Research Fellow, Russell Mannion is a Research Fellow and Peter C. Smith is Professor, at the Centre for Health Economics, University of York.

ACKNOWLEDGEMENTS

FURTHER COPIES

Further copies of this document are available (at price £5.00 to cover the cost of publication, postage and packing) from:

The Publications Office
Centre for Health Economics
University of York
York YO10 5DD

Please make cheques payable to the University of York. Details of other papers can be obtained from the same address or telephone York (01904) 433648 or 433718.

ABSTRACT

The provision of quantitative information has been given a key role in securing good performance in the new NHS. A new National Performance Framework has been proposed encompassing a number of dimensions of performance. Whilst this approach to managing the NHS is welcomed, it is essential to understand the strengths and limitations intrinsic to the use of performance indicators for this purpose. In particular, complex behavioural consequences may arise in response to the collection and dissemination of performance data, some of which may be unintended, potentially dysfunctional and damaging for the NHS. Results from a recent study on the performance of NHS Trusts are used to illustrate the sort of unintended side-effects which occur within the current system and which may in principle be replicated in the new system in future. Whilst the possibility of such consequences does not invalidate the potential of the new Performance Framework to secure the desired changes in the NHS, it does suggest that careful attention needs to be paid to the assessment of unanticipated side-effects.

(I) INTRODUCTION

The publication of the 1997 White Paper, "The New NHS: modern, dependable" (and its Scottish and Welsh equivalents) signal a fundamental change in the way that the NHS is to be managed (Department of Health 1997). The new NHS is to be "based on partnership and driven by performance". The government's 'third way' is neither an attempt to return to the centralised command and control systems of the 1970s, nor a commitment to retaining the competitive internal market system of the 1990s. The stated intention is to build upon the strengths of each, creating a more collaborative service with economic relationships based on partnership and co-operation. Whereas the internal market reforms gave competition the central role in securing performance, the key instrument for managing the NHS in the future will be the provision of quantitative information, in the form of a new National Performance Framework. Proposals for the nature and content of the Framework were issued for consultation in January 1988, providing more detail on the broad intentions outlined in the White Paper (NHS Executive 1998).

In general, the new approach to managing the NHS is to be welcomed. However, in developing the Performance Framework, it is essential to have a proper understanding of the strengths and limitations intrinsic to the use of performance indicators for securing management control in the NHS. A key issue which has so far been largely ignored in the White Paper and the consultation document is the impact of performance indicators on behaviour. If it is to have any effect, publication of performance data must lead to changes in behaviour of NHS organisations and staff. Yet there is little appreciation in the official documents of the complex behavioural consequences that may arise as a result of the collection and dissemination of performance data. In addition to having the predicted impact on behaviour, the potential exists for performance management systems to produce unintended and possibly dysfunctional consequences which may be damaging to the NHS.

This paper summarises the theoretical issues and the results of previous research in this topic. It then uses some of the findings from a recently completed study to illustrate the sort of unintended side-effects that occur within the current system of performance management which may in principle be replicated in the new system in future. We conclude that careful attention needs to be paid to the potential for unanticipated side-effects. The possibility of such consequences does not invalidate the principle of seeking to develop a workable Performance Framework. However it does suggest that much more attention than is evident in the consultation document must be directed at how the proposed data are to be disseminated and used within the NHS.

The paper is arranged as follows. The next section provides an overview of the new performance framework as set out in the White Paper. Section III sets out a theoretical model of the performance measurement process and Section IV summarises the key research findings from attempts to introduce comparable measurement schemes in the public sector. This is followed by a brief outline of the aims and methodology of the Department of Health Funded study in the information used to assess NHS Trusts. Section VI summarises some of the findings of this study relating to the unintended and dysfunctional outcomes of performance indicators and makes some suggestions for a range of strategies to offset some of the unanticipated consequences described. Finally, in Section VII the paper outlines a number of policy recommendations relating to the future development and implementation of the national Performance Framework.

(II) THE NATIONAL PERFORMANCE FRAMEWORK

The White Paper and the consultation document outline the six key areas of performance on which the National Framework will focus:

- **Health improvement**

The aim of improving the general health of the population.

- **Fair access**

Fair access to health services according to need irrespective of geography, class, ethnicity, age or sex.

- **Effective delivery of appropriate healthcare**

The aim of providing effective, appropriate and timely health care which meets agreed standards.

- **Efficiency**

The use of NHS resources to achieve best value for money.

- **Patient/carer experience**

The way in which patients and carers view the quality of treatment and care that they receive.

- **Health outcomes of NHS care**

Assessing the direct contribution of NHS care to improvements in overall health.

Targets for progress against these six areas are to be built into: the performance agreements between NHS Executive Regional Offices and Health Authorities; local Health Improvement Programmes; accountability agreements between Primary Care Groups and Health Authorities; and the agreements between Primary Care Groups and Trusts. The measures used to assess each of these dimensions will be constructed in the light of responses to the consultation document. It is expected that the Performance Framework will replace the Purchaser Efficiency Index in 1999. The proposals for assessing and managing the performance of the NHS are to be welcomed on a number of grounds. Specifically, they represent an improvement on existing arrangements in the following areas:

- Focus on the health aspects of performance rather than only on finance and activity issues
- Focus on equity issues
- Focus on patient experience

- Acknowledge the contribution of other agencies besides the NHS to health outcomes
- Focus on the population or client group rather than solely on providers
- Focus on the local as well as the national level

More generally, the increased emphasis to be placed upon performance measurement in the NHS is to be welcomed. Excessive reliance on markets to secure improvements in health services is misguided. Successful performance management is likely to require a far more sensitive set of instruments than a competitive market, and a good system of performance measurement must form an essential component of that set.

(III) THEORETICAL CONTEXT

It is probably simplest to think of the performance measurement approach as one of *feedback*. Figure 1 outlines a very simple model encapsulating the essence of this process. Certain aspects of the organisation are measured. These are analysed and interpreted in the light of environmental influences and organisational objectives. Finally, some sort of action results which alters the nature of the organisation. The process then continues.

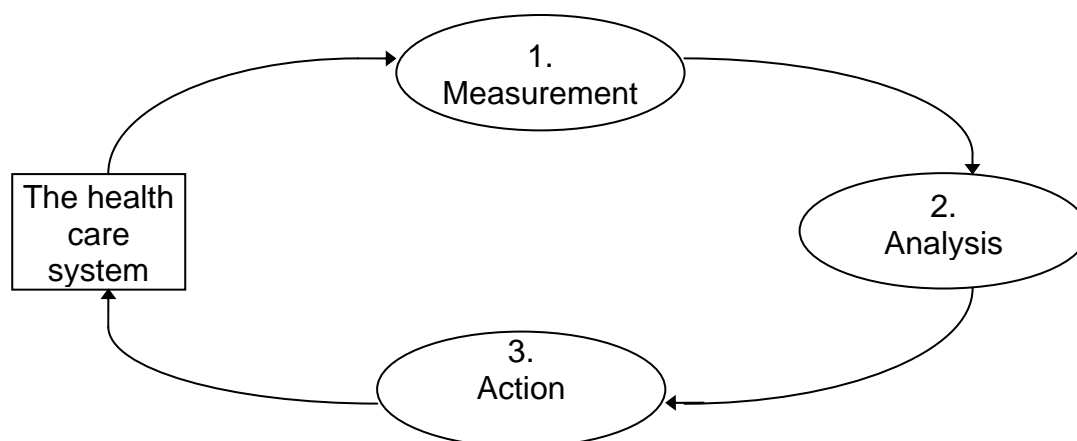


Figure 1: A Model of the Performance Measurement Process

Many of the consequences of this system of feedback are likely to be intended and helpful. For example, publication of data relating to waiting times for elective surgery, in association with substantial political attention to the issue, has undoubtedly contributed towards an improvement in certain aspects of waiting within the NHS. Less certainty however, can be attached to the broader consequences of such measurement. For example, at various times it has been claimed that the emphasis on very long waits has led to increased emphasis within NHS Trusts on relatively minor procedures at the expense of more urgent surgery, and that the emphasis on elective surgery has resulted in the diversion of resources from activities for which equivalent performance measures do not exist.

Thus the net effect of any single performance measure (or set of measures) on the NHS as a whole may be quite complex to evaluate. In particular, it is clearly important to look well beyond any improvements in the measure itself in order to determine whether publication of the measure is on balance beneficial. In this section we therefore summarise a variety of ways in which unintended side-effects can arise from performance management systems. The discussion considers these effects under the three headings: measurement, analysis and action.

Measurement stage

Any measurement system will to some extent be *partial*. If it is intended to affect the behaviour of managers, clinicians and other workers, it will require incentives to pay increased attention to the measured, and therefore conversely less attention to the unmeasured. This may induce tunnel vision, or the neglect of unmeasured aspects of activity.

A slightly different danger is that of *fragmentation*. Many outcomes from health care depend on the collaboration of a variety of other organisations, both within and outside the NHS. The eventual outcome therefore depends on the joint efforts of these agencies. It may be quite challenging to devise measurement schemes which can capture the contribution of each agency to outcome, and there is a danger of sub-optimisation, in the sense that agencies are driven towards pursuing their own performance criteria with little regard for joint outcomes.

Almost all measures are *short term* in their perspective. Yet current measures may reflect the outcome of years' of health care activity, and therefore cannot be attributed solely to current activity. And current activity may affect future outcomes, which are not captured in current performance measures. Special attention will therefore have to be given to activities which have long term health outcomes.

Although there may be universal agreement that a particular phenomenon is an important aspect of NHS performance, there may be numerous *instruments* for capturing the phenomenon of interest. For example, existing measures of waiting time choose to focus on exceptionally long waits, but there is no reason why this particular aspect of waiting should necessarily be addressed. Yet the choice of one instrument in preference to another may have profound implications for organisational behaviour. There is a danger of encouraging *measure fixation* if managers pursue strategies which enhance their success in terms of the reported measure, rather than focusing on the underlying objective.

It must be remembered that the collection of many indicators of performance will rely on the co-operation of front line NHS staff. If their rewards are to any extent dependent on the data they report, there is of course a constant danger of *misrepresentation* in one form or another.

Finally, it is important to bear in mind that health care is a dynamic, ever-changing activity, and that technologies and needs change very quickly. There is therefore a danger of *ossification*, and any performance measurement scheme therefore needs to be constantly reviewed to ensure that it captures current priorities.

Analysis stage

Health care performance data are often immensely difficult to interpret. Many measures reflect the outcome of a complex interaction of factors, such as environment and deprivation, which may have a profound impact on the measure which is outside the control of the NHS. Careful analysis may therefore be required to ensure that the *contribution* of the NHS to the chosen performance measures is correctly identified. In practice, the analysis of performance data is in its infancy, and the potential for faulty inferences may be large.

Action stage

The impact of a performance measurement scheme will depend on the rewards, punishments and incentives implicit in its design. At one extreme, if no appropriate incentives are put in place, the data produced may be largely ignored and produce no meaningful action. At the other extreme, the livelihood of managers and health care professionals may depend crucially on reported measures, leading to the potential for excessive attention to reported performance (as opposed to patient outcome) and misrepresentation.

Furthermore, certain types of measure may be susceptible to *gaming* on the part of health care professionals. This phenomenon was endemic to the Soviet Union, where year on year improvement was the usual performance measure. Managers had little incentive to achieve good results in the current year because in all probability their future targets would be based on the new level of attainment, leading to the need for increased effort in the future.

(IV) PREVIOUS RESEARCH ON BEHAVIOURAL RESPONSES

This section summarises the limited amount of research evidence available in relation to the behavioural impact of performance data publication. A more detailed account can be found in Nutley and Smith (1998).

Publication of performance data has little purpose if it is not to have some impact on the behaviour of the health care organisation. The nature of the feedback that organisations receive from interested parties following publication is therefore a key influence on subsequent managerial and clinical actions. Such feedback might be formal, as in the case of the Peer Review Organisations advocated by the US Health Care Financing Administration (HCFA), which are intended to offer "non-punitive" interpretation of results. Or the feedback might be more haphazard and informal, as for example in the Patient's Charter initiative. Here data are exposed in the public arena, from which a variety of political pressures might emerge. Indeed the implications of feedback may be very different depending on whether or not data are made public. For example, Conway et al (1995) describe a feedback system in Maine in which data are shared in confidence with clinicians in order to retain their support for the project.

As noted above, intrinsic to the feedback process is the system of punishments and rewards attached to the performance measurement scheme. These are likely to be largely implicit and difficult to define, but may nevertheless be of immense importance. In concrete terms, performance rankings might affect budgets, bonuses and job security, while less tangibly they are likely to influence staff morale and recruitment. In the US a surgeon's poor league table rating might lead to loss of livelihood. Moreover, a potential problem in the feedback process is the existence of contradictory messages. For example, the US Health Quality Improvement Initiative emphasises consensus, collaboration and non-antagonistic feedback. Yet Nash (1995) points out that HCFA are also basing accreditation and payment decisions on performance data, a conflict which may undermine many of the benefits of feedback.

While undoubtedly of central importance, the nature of the feedback process and the associated reward structure are intrinsically difficult to study, and there has been a notable lack of research on the impact of performance publication within the health care sector (Rosenthal and Harper 1994, Nash 1995). Any user of health care performance data is likely to need supporting material to help interpret the results. This might take the form of supplementary data or of expert help. For example, Goldman and Thomas (1994) describe how hospital specific mortality rates provided a screening tool which led onto a focused medical records review in outlier providers; Nelson et al (1995) report the use of supporting data in the form of "instrument panels"; Finlan and Zibrat (1994) describe the impact on clinical practice in one hospital of comparative data provided by the US Joint Commission on Accreditation of Health Care Organizations; and Moller *et al* (1994) describe how a consortium of providers have set up a study group to examine outcome measures in paediatric cardiac care. However, there has in general been little examination of the mechanism for feeding information back to hospitals and clinicians. Most schemes appear to rely on a vague hope that providers will "do something" in response to the data. Yet evidence from the broader UK public sector suggests that - in the absence of formal feedback procedures - little change may occur.

One of the key concerns about many performance measurement systems, and particularly report cards, is that they are backward looking, judgmental and poor motivational devices. Greene and Wintfeld (1996) conclude that by the time the data are published, users of the report cards can have little confidence that the ratings are still applicable. As a response to these concerns, some authors like to think of performance indicators in terms of the instrument panel of an aircraft or car (Jackson 1995, Meekings 1995). The focus then is on providing real time, on-line monitoring which is action oriented. Although Nelson *et al* (1995) illustrate the instrument panel concept by reporting on work in the Dartmouth-Hitchcock (US) health care system, it is not generally clear how these principles can be translated to the health care setting.

More generally, few studies have sought to evaluate in any comprehensive way the behavioural ramifications of publication. The obvious means of judging the impact of a performance measurement scheme is by examining its apparent impact on performance. Such exercises may however be tautological, in the sense that the yardsticks used to evaluate the scheme may be precisely the same potentially imperfect instruments of measurement used in the scheme itself, which may be subject to the problems of partiality, distortion and incompleteness we have noted above. Thus, the previous UK Government proclaimed the success of the Patient's Charter by citing improvements in the length of time patients wait for inpatient admission, the Charter's prime performance measure (UK Government 1995). Yet there are many other aspects of NHS health care on which the Charter is silent, but which may have been adversely affected by the emphasis on waiting times.

In the same way, Chassin, Hannan and DeBuono (1996) cite a notable decline in risk-adjusted mortality associated with the New York report card scheme. However, Greene and Wintfeld (1996) argue that the phenomenon may be more a function of an increase in the reporting of risk factors than a real improvement in surgical outcomes. And Ghali *et al* (1997) note that similar improvements in risk-adjusted mortality were found in Massachusetts, which did not publish mortality data. Naive examination of apparent improvements in reported performance is therefore inadequate, and much more complete evaluation, seeking to identify any side effects of performance publication is needed. In this context, Smith (1993) examined the impact on maternity managers of a single, well-established performance measure - the perinatal mortality rate. He identified seven potentially dysfunctional consequences of publication of these data, and found compelling evidence that some had indeed materialised. For example: managers were concentrating on perinatal mortality at the expense of other dimensions of performance; maternity units were paying less attention than previously to the effects their actions might have on other parts of the health service; long term outcomes were being given less attention than hitherto; and there was some suggestion that doctors' reporting of the timing of mortality may have been influenced by the definition of perinatal mortality (which excludes the first 28 weeks of pregnancy).

Thus the feedback phase of the performance measurement process clearly requires more attention. As Stewart (1984) points out for the broader public sector, in examining the accountability relationship it is important to consider both the giving of an account (the publication of performance data) and the mechanisms available to hold organisations to

account. The accounts provided to stakeholders within health care systems need to meet their varied interests and be communicated in a language which makes sense to them. There is then a need for mechanisms with which to hold service providers to account. The strengthened role of external audit bodies is one means of achieving the latter. Where a market exists for health care services, it may be possible for patients and purchasers to vote with their feet and choose providers on the basis of reported performance. However within a publicly provided system, there is also a clear need for transparent democratic processes for holding providers to account.

In summary, whatever the system of health care, simply reporting performance is unlikely to be sufficient to secure improvements.

(V) AIMS AND METHODS OF STUDY

The aim of the research was to investigate the types of information used to assess the performance of NHS Trusts. A major element in the research was the exploration of the kinds of 'soft' or informal information used alongside more formal measures of performance. The results of this will be reported elsewhere. However, we also sought to identify any unintended and dysfunctional behavioural consequences of current performance measures. Although the study was based on experiences with the current set of performance measures, it is of relevance to the new framework for two reasons. First, many of the indicators proposed will be the same as (or very similar to) those used currently (eg waiting times); second, the principles relating to behavioural responses will apply to a range of measures which have characteristics similar to those considered, even if the specific measures differ.

The research was based on case-studies of eight Trusts. Two NHS Executive Regional Offices were approached and senior staff responsible for provider finance and performance were asked to nominate four hospitals exhibiting what they considered to be a wide range of performance. We limited the choice to District General Hospitals, in order to make our sample as homogenous as possible. The Chief Executive of each of the selected Trusts was contacted and invited to participate in the research.

Within each Trust semi-structured interviews were undertaken with the Chief Executive, the Medical Director, a Nurse Manager and a Junior Doctor.¹ One Junior Doctor did not attend for the interview, so we undertook a total of 31 interviews within Trusts. In addition, we interviewed the Finance Director of the local Health Authority for each Trust (the Trust's main purchaser). We asked the finance director to include any other relevant staff in the interviews and in half the cases additional people attended, mainly from a health development or commissioning background. At the Regional Offices, we interviewed staff responsible for provider finance and performance. These external organisations added a further 10 interviews to the dataset and thus in total the study is based on the results from 41 interviews.

¹ One of the aims of the study was to investigate the views of people at different levels in the organisation on how their performance was judged and how they felt they were held to account. However, the results of this are not discussed further in this paper but will be reported elsewhere in the future. We also gathered views of a sample of GP fundholders, but again, these are not discussed in this paper.

(VI) THE UNINTENDED CONSEQUENCES OF PERFORMANCE INDICATORS

Following Smith (1995), we report the findings on unintended consequences of performance indicator schemes under the following headings:

Tunnel vision: Concentration on areas that are included in the performance indicator scheme, to the exclusion of other important areas.

Sub-optimisation: The pursuit of narrow local objectives by managers, at the expense of the objectives of the organisation as a whole.

Myopia: Concentration on short term issues, to the exclusion of long term criteria that may only show up in performance measures in many years' time.

Misrepresentation: The deliberate manipulation of data including 'creative' accounting and fraud so that reported behaviour differs from actual behaviour.

Gaming: Altering behaviour so as to obtain strategic advantage.

We consider these below in turn, taking into account the range of views received from different organisations and from people at different levels within the Trusts.

Tunnel Vision

There was clear recognition by staff at all levels that current indicators did not give a 'rounded' view of the performance of a Trust and that their specific focus often diverted attention from equally legitimate (but unmeasured) aspects of Trust performance:

"If you set 10 performance indicators what this will do is encourage the Trust to perform on these 10 and not be too bothered about anything else" (Health Authority Finance Director).

"It (the current system of performance indicators) cannot possibly give a rounded view because they are only a handful relative to the whole spectrum that goes on in a Trust" (Trust Medical Director)

"There is no point in only trying to manage what you can measure. You have to be imaginative about other things and work out ways of trying to assess them" (Trust Chief Executive)

The current priority given to waiting times targets was frequently cited as diverting attention and resources away from other important spheres of Trust performance. Indeed it is apparent that a vast amount of effort and resources are devoted by Health Authorities and Trusts to ensuring they meet the targets and some clearly feel it is money which they would prefer to spend elsewhere. In some cases, those responsible for ensuring the target is met will use a range of tactics to ensure they achieve the correct results, sometimes at the expense of other priorities.

Examples of both the time and effort involved and the sort of tactics employed at Trusts to ensure targets are met included:

- a senior person in one Health Authority spends the majority of their time chasing the Trust about individual patients on the waiting list who have to be treated in order to avoid a breach of the standards
- treating all the cataract patients on the list in order to reduce the numbers at a relatively modest cost, whilst leaving people requiring more expensive operations on the list
- removing long waiters, regardless of their medical condition
- putting people on the "pending list" (a waiting list for the waiting list) rather than on the official waiting list
- the employment of 'hello nurses' in A and E Departments in order to ensure the 5 minute waiting time target is met

In such cases, staff feel they have been pressured to "manage the waiting lists rather than the patients" and that changes have been made "not to improve the quality of care but to improve our ratings" (Trust Medical Director).

The propensity for tunnel vision may be associated with the particular management style of a Trust:

"We are not performance indicator driven but I know that you could go to others and it would all be about performance indicators. The Trust we are merging with couldn't be more different in that respect. They would say.....if you cannot graph it or put a number to it, it doesn't exist"

Clearly, the introduction of a wider range of measures, including clinical indicators will offset to some extent the common complaint that the current system leads managers to focus only on the process orientated measures, at the expense of more important things which may not be measured (or measurable).

The lack of measures relating to outcomes, compared with the number focusing on process and activity, was mentioned frequently in our study. The piloting of clinical indicators and the proposals for the national performance framework were therefore welcomed by most staff. However, difficulties in finding good comparators and methods for dealing with case-mix were seen as potential problem areas. Given the perceived limitations, some clinicians advocated that clinical outcome measures should be supplemented by process measures of clinical activity such as the recording of "near misses". The use of evidence based guidelines and "clinical pathways" were also thought to be important in augmenting outcome measures.

Adding new measures may therefore reduce the extent to which people focus only on demonstrating success on a small number of measures, but too *many* measures may also induce dysfunctional consequences, particularly if it becomes a 'juggling' act to try to focus on them all. For example, when asked about the new proposals for performance management, one Health Authority Finance Director reported that even without the addition

of the new indicators he felt there were too many. No-one could be expected to do well on them all and thus it would result in a considerable waste of time and effort as people tried to discover what the "real" key indicator was.

Sub-optimisation

Sub-optimisation can occur where there is a lack of congruence between personal incentives and the global objectives of the organisation. This was one of the key aspects of performance which we addressed in our study and we will be reporting on this in detail elsewhere in future. However, in brief, we found that although each Trust had corporate objectives, these objectives were not always aligned with the specific incentive structures for different staff within the organisation. In particular, it was clear that it was often difficult to align the Trust financial objectives with specific clinical priorities. In order to address this problem, all of the Trusts were attempting to actively engage medical staff within the management process to facilitate clinical 'ownership' of the major strategic financial and administrative issues facing the organisation. This was thought by some respondents to be a particularly difficult task when a Trust was in a serious financial position and clinicians were expected to co-operate in service reductions:

"The Trust objectives as a whole are business type objectives. What I'm trying to do now is get enough information to turn them into more clinical objectives... . What I want [is for the clinicians] to own the decisions and own the outcomes" (Trust Chief Executive)

"One of the key reasons why this organisation has failed is because the clinical staff, junior and senior have largely been disengaged from the management process. So part of the work we have been doing to address our financial deficit has not just been to take a substantial amount of money out of the cost expenditure it has also been to bring clinicians into the heart of the management process." (Trust Chief Executive)

Interviews with the junior doctors also revealed a lack of congruence between their own personal objectives - which were largely related to getting a good reference for their next job - and those of the Trust in general. Indeed, many of them reported little detailed knowledge of the formal performance management system at all:

"I don't have much to do with performance indicators. They don't impact on my work. I haven't seen the [clinical indicators] at all" (Junior Doctor)

Although there were some exceptions where junior doctors did realise they had to see patients within a certain length of time in order to comply with Patient's Charter demands, most saw themselves as being distant from the formal systems of performance assessment. Another example of sub-optimisation identified by senior staff was the potential for a lack of congruence between the incentives or objectives of different agencies responsible for provision of health care within a local community. The most commonly cited example of this was "bed blocking" which can occur as a result of lack of agreement about responsibilities of health and social services.

Aside from trying to promote ownership of objectives by clinical staff by involving them in the setting of targets, other strategies for dealing with sub-optimisation were discerned. We

found that setting devolved budgets and undertaking individual performance review or formal appraisal schemes were the two main mechanisms for transmitting Trust level objectives down to staff at various levels within the organisational hierarchy.

Additionally, some Trusts had used comparisons with clinical services in other Trusts to combat professional challenge of corporate objectives:

"We've only been able to break into the professional dominance by challenging such things as staffing levels based on benchmarking at other Trusts..... We've also maintained a professional nursing support group where senior nurses can express their views about skill mix etc." (Chief Executive)

Myopia

There was a general feeling among our respondents that many of the current performance indicators are short term in nature and actions taken now may not show up in indicators for several years:

"I think short-termism has been a major problem in performance management culture of the NHS eg the fact that there has been until recently the requirement to make income and expenditure balance each year. Now we have failed to do that and we have been treated like lepers for doing so. But when you are going through major change any private sector enterprise would not just look at one years results but achievements against a five year plan" (Trust Chief Executive)

In fact, most of the respondents in our sample were fairly optimistic about this - although they felt that they were sometimes being pushed to deliver short term targets (e.g. reductions in management costs) without a view to the longer term, there was some recognition that sometimes the short term picture will be poor but that this was acceptable if it produced a better long term outcome. For example, Regional Office staff recognised that some Trusts had performed poorly in terms of their financial duties but that this was not a concern as it was to be addressed through the longer term financial strategy. As long as the use of supplementary and softer information remains as a complement to the hard performance indicator data, this may not become a significant problem in the future.

The perspective of those people responsible for taking actions which influence long term outcomes will be influenced by the likelihood that they will still be in post when the outcomes are revealed. For example, some of the Trust Chief Executives in our study had inherited a very poor financial or clinical situation from the previous post holder and it may be the case that some of the actions taken previously would not have been taken if the post holders had anticipated they would have been around to face the consequences. As one Chief Executive who had recently taken over a poorly performing Trust stated of his predecessor:

"The last guy got out of here before it hit the fan. Timing is all"

Longer employment contracts may encourage behaviour which takes account of the impact of current actions on future targets. Alternatively, ensuring that at least some reliable and

appropriate process measures are included in a performance management system would provide an early signal of whether behaviour was likely to produce the desired long term outcomes in the future.

Misrepresentation

Misrepresentation is the deliberate manipulation of data so that reported behaviour differs from actual behaviour. The scope for misrepresentation of data within a Trust environment is particularly broad because many of the data used to measure performance and hold staff to account are under the direct control of those staff. A mixed picture emerged in our study, with some people citing specific instances of misrepresentation (either by them personally or by others), whilst others said they were confident this did not happen within their organisation. Examples of creative measurement cited included:

- double counting of FCEs when a patient is referred to another consultant within the same hospital
- excluding the least favourable when calculating results e.g. not including out-patient clinics where a consultant was sick or was called to casualty so that the 30 minute waits are achieved
- changing the order in which the codes for procedures undertaken on patients are entered. For example, the pressure to achieve a low rate for D&Cs can encourage staff to enter the D&C as the second, rather than the primary procedure undertaken on a patient because only the first code is included in the returns.

Some of this manipulation is undertaken because people are keen to give the impression they are meeting the targets and the degree of flexibility involved in recording and reporting the data makes this possible, without them having to resort to outright lies. Most people said they felt they were "massaging" data rather than misrepresenting it and that all they were doing was presenting themselves in the "best possible light". There was some evidence that Health Authorities and Trusts sometimes collude in this, especially when the NHS Executive is taking a tough line with failures to meet targets.

"... you look for ways that you can conspire with the Health Authority to give them what they want to hear up there. In a way it minimises your pain and maximises their gain". (Health Authority Finance Director)

"Yes we definitely did that [collude with the local Trust to misrepresent efficiency index figures]. The only way that Health Authorities were gaining resources was to show that they were increasing care in areas that they were not. But what we felt was that once you start fiddling the figures you have to keep a consistent message year on year and it gets further away from reality" (Health Authority Finance Director)

Thus there is an incentive to just ensure the information is reported in such a way as to give those on the ground a 'quiet life'. Similarly, some Health Authorities noted that although they were pleased that their local Trusts met the Patient's Charter standards, they could not help wondering whether this was an accurate reflection of what actually went on at ground level or whether the Trusts were skilled at ensuring the data were recorded in a way that gave the right results:

"..... they are very good at demonstrating they are very good at meeting the Patient's Charter, but that doesn't mean they have met the spirit of what you are trying to achieve". (Health Authority Finance Director)

Paradoxically, because there has been a tendency to provide additional finance for those failing to meet the waiting times targets, there is also sometimes an incentive for Trusts to misrepresent their data in the opposite direction, extending their waiting times in order to obtain additional funds. Some Trust Chief Executives noted that it was relatively easy to "extend" their waiting lists in order to provide a good argument for a share of extra funding (especially central funds).

Although many examples of misrepresentation were supplied, others in our sample stated that even though they know they could distort the information, they chose not to, either as they felt it was dishonest or because they feared detection. For example, one Medical Director said he had been approached by a casualty consultant about reducing the number of people who were sent home rather than admitted as - if they admitted them - the patients would count as extra FCEs, hence boosting their activity data. The Medical Director declined on the grounds that this was poor clinical practice and that they would bound to be detected. Others felt it was wrong to cover up a problem by fiddling the figures as the root cause of the problem would then never be tackled. In addition, many of the junior medical staff were not even aware of what the current measures are so they were certainly not in a position to undertake a great deal of data manipulation.

Misrepresentation of data can also arise when those responsible for reporting do not place sufficient importance on it, because they feel the measures are not valid and are not therefore willing to devote effort to ensuring the data are valid. A particular example of this was revealed in our interviews when a Medical Director admitted that he had "scandalised" his junior colleagues by suggesting that:

"..... we save ourselves 5 hours and make the figures up".

In this case, the indicator for patients' complaints required the Trust to use a different classification system for the returns to the NHS Executive than the system they were using for returns to the purchaser. The Medical Director felt justified in having a guess rather than going through a bureaucratic process as he did not feel they were doing it for gain (other than saving time).

An additional aspect of data misrepresentation which does not involve the deliberate manipulation of data relates to the general quality of the data used to compile the measures. A number of respondents mentioned the "garbage in - garbage out" issue whereby they felt that although the measures may be acceptable in principle, in practice the data used to compile them were so poor and subject to so many errors or distortions, that they felt no confidence in the figures that come out at the end. For example, in relation to the proposals for national reference cost schedules, a number of purchasers remarked that they had already tried cost benchmarking and had identified significant problems using it just on a local basis. One problem is caused by the difficulties in ensuring the cost data used by Trusts are

comparable and measure the same thing (even if it is done on a HRG basis and is supposed to follow central guidance on methodology), the results of the exercises raise more questions than answers. In the words of one Health Authority Director of Finance,

"there are clouds and clouds of smokescreens around the differences - you cannot just say, you [the Trust] are twice as expensive as the other".

A similar point was made in relation to the waiting times data by another Health Authority and also in relation to the discrepancies in what is actually counted as a day case or an inpatient stay in some of the Trusts. It was clear that without a lot of effort on the part of those monitoring the indicators, the results themselves are not very useful when the quality of the data used to produce them is questionable. Many of the indicators proposed in the new framework will rely on data which are similarly subject to a degree of interpretation or are of variable quality.

One possible way of reducing the extent of misrepresentation is to increase audit activity in order to detect it and to penalise those responsible. However, the costs of any increased monitoring and audit would have to be weighed up against the benefits. Providing incentives to record data honestly and consistently may also reduce misrepresentation, although it was clear from our study that it was not always obvious where the "blame" for this behaviour would rest. For instance, although the doctors or nurses might record information truthfully, the data could be "massaged" further up the management line by those who gain from apparent good performance. Conversely, even if senior staff discourage misrepresentation, if those responsible for providing the data have no faith in the measures, the incentives may not be sufficient to motivate them to ensure the accuracy of the data is maintained. The key is to find a balance by ensuring that those responsible for collecting, recording and reporting the data (which includes the most junior staff as well as the senior Executives who are held to account for performance) believe the measure is reasonable and worthwhile.

Gaming

The most frequently cited example of gaming concerned the efficiency index. Respondents indicated they would be reluctant to achieve high gains one year for fear that they would be expected to deliver the same or higher gains in the future. This ratchet effect was a common feature of the former Soviet economic system where managers were 'punished' for good performance by having higher standards set in the subsequent years plan. Where this occurs managers have little incentive to increase productive efficiency and reduce organisational slack. In our study, there was some acknowledgement that all parties knew that gaming went on and acted accordingly. Frequently an analogy to participating in a "game of chess" was made. One Trust in the study which started out performing very well on the efficiency index tables, saw that there was little scope for matching these achievements in the following years so decided to reduce performance in certain areas so that they could get credit for improving performance at a later stage:

"In the second and third year (of the efficiency index) we began to realise that simply working our staff harder and keeping the wage bill down was not necessarily the right way to go. So we then had a plan of sliding down the table and accepting that our position on the table would get worse-because everyone wants to get better and slide up the table" (Trust Medical Director)

However, some Trusts pointed out that it had become easier over time to reach an agreement with the Health Authorities about the need to moderate future demand for efficiency gains if they had managed to achieve high gains in a particular year (especially if this was due to a specific service re-configuration which would not be repeated the following year). Some felt that the game playing wasted a lot of time and effort and that they preferred to avoid it if they could. Although the efficiency index is due to be replaced, it is worth noting the problems above as they will relate to any system whereby the current performance target is based on past performance.

The financial regime in which Trusts operate was also felt to encourage some gaming. It was noted that Trusts which fail to meet their financial duties will often get "bailed out" by the region and thus there may be an incentive to fail to meet these targets - not in a spectacular way which would attract a great deal of attention, but sufficient to allow overspending without penalty. One Trust used the example of a neighbouring Trust which had managed to get five star ratings on the Patient's Charter targets by overspending and will not be required to meet their deficits the following year. Similarly, those who fail to meet waiting times targets are more likely to receive a share of the extra waiting list funds made available from central resources than those who have already achieved the targets.

One strategy for minimising gaming is to use a range of performance measures rather than just one which may be easily gamed. Although there are currently a range of indicators for the NHS, a great deal of emphasis has been placed on achievement of the efficiency target and as this is also sometimes linked to explicit rewards for senior staff, it is therefore not surprising that gaming has focused on this specific indicator. Although the efficiency index is to disappear, it is important to ensure that the reward structure for meeting any targets is flexible enough to respond to managerial efforts which contribute towards improvements in future targets as well as the achievement of current targets. Development of benchmarks of performance which are independent of the organisation's past behaviour can also help to minimise the gains from gaming. Although comparative analysis of performance brings a different set of problems, if managers understand that their future targets will not be influenced to any great extent by current performance, they are more likely to seek out opportunities to improve their performance.

(VII) SUMMARY AND POLICY RECOMMENDATIONS

In this paper we have outlined the importance of ensuring that attention is paid to the potential unintended and dysfunctional consequences induced by any system of performance measurement. It is clear that performance measures do not always effect the desired changes in behaviour. The unintended consequences of some measures may actually encourage people to behave in ways which are directly contradictory to what was expected. Although respondents were referring to the current indicators which may be changed in the future, many will be similar to (or in some cases, the same as) existing indicators. Moreover, there is no reason to expect substantially different behavioural responses even for completely new indicators.

Techniques to mitigate some of the specific unintended consequences highlighted by respondents have been suggested and it is also possible to identify some general strategies which address the dysfunctional outcomes, as noted by Smith(1995). The first four strategies address a large number of problems, and so are likely to be applicable in most situations:

- Involving staff at all levels in the development and implementation of performance measurement schemes;
- Retaining flexibility in the use of performance indicators, and not relying on them exclusively for control purposes;
- Seeking to quantify every objective, however elusive;
- Keeping the performance measurement system under constant review.

The importance of the next three strategies is more dependent on the particular aspect of performance being measured, being most relevant when objectives are poorly defined and measurement of output problematic:

- Measuring client satisfaction;
- Seeking expert interpretation of the performance indicator scheme;
- Maintaining careful audit of the data.

The final three strategies are designed to address specific difficulties - myopia, misinterpretation and gaming - and so should be considered when any of these is especially important. Note however that they may have negative effects relevant to other dysfunctional phenomena as highlighted earlier.

- Nurturing long term career perspectives amongst staff;
- Keeping the number of indicators small;
- Developing performance benchmarks independent of past activity.

The extent to which these and other potential solutions are relevant for the new National Performance Framework is a matter for further research and consideration.

In conclusion, we advocate a cautious approach to the implementation of the National Performance Framework. Current proposals represent an improvement over the previous methods of managing the NHS, but it would be a mistake to assume that only positive results will be produced. We have illustrated the potential for unintended and dysfunctional side-effects in the NHS and we have suggested some strategies which may be used to help reduce the extent to which they arise in the future.

However, until more is known about the behavioural impact of performance management systems, there will always be a danger that new systems will perpetuate the tendencies towards dysfunctional behaviour and even create additional ones. This suggests that the new National Framework should be evaluated not only in terms of the degree to which it produces the expected improvements in the chosen measures, but also in terms of the unanticipated side-effects. A major research task is to identify the characteristics of schemes which influence successfully the behaviour of NHS staff. This would involve consideration of the impact of alternative dissemination methods (eg league tables, target setting, benchmarking, informal review) as well as the incentives and sanctions associated with the indicators (eg financial, peer pressure, reputation). This is the subject of our future research agenda.

REFERENCES

- Chassin, M.R., Hannan, E.L. and DeBuono, B.A. Benefits and hazards of reporting medical outcomes publicly. *New England Journal of Medicine* 1996; **334**(6): 394-398.
- Conway, A.C., Keller, R.B. and Wennberg, D.E. Partnering with physicians to achieve quality improvement. *The Joint Commission Journal on Quality Improvement* 1995; **21**(11): 619-626.
- Department of Health *The New NHS: Modern, Dependable*. HMSO, London, 1997.
- Finlan, J.K. and Zibrat, F.S. Struggles and successes: experiences of a beta test site. *The Joint Commission Journal on Quality Improvement* 1994; **20**(2): 49-56.
- Ghali, W.A., Ash, A.S., Hall, R.E. and Moskowitz, M.A. Statewide quality improvement initiatives and mortality after cardiac surgery. *Journal of the American Medical Association* 1997; **277**(5): 379-382.
- Goldman, R.L. and Thomas, T.L. Using mortality rates as a screening tool: the experience of the Department of Veterans Affairs. *The Joint Commission Journal on Quality Improvement* 1994; **20**(9): 511-521.
- Greene, J. and Wintfield, N. Report cards on cardiac surgeons: assessing New York State's approach. *New England Journal of Medicine* 1996; **332**(18): 1229-1232.
- Jackson, P. (ed) *Measures for success in the public sector*, London: Public Finance Foundation, 1995.
- Jordan, H.S., Straus, J.H. and Bailit, M.H. Reporting and using health plan performance information in Massachusetts. *The Joint Commission Journal on Quality Improvement* 1995; **1**(4): 167-177.
- Meekings, A Unlocking the potential of performance measurement: a practical implementation guide *Public Money and Management* 1995; **15**(4): 5-12.
- Moller, J.H., Powell, C.B., Joransen, J.A. and Borbas, C. The paediatric cardiac care consortium - revisited. *The Joint Commission Journal on Quality Improvement* 1994; **20**(12): 661-668.
- Nash, D.B. Is the quality cart before the horse?. *Journal of the American Medical Association* 1992; **268**(7): 917-918.
- Nash, D.B. Quality of measurement or quality of medicine? *Journal of the American Medical Association* 1995; **273**(19): 1537-1538.
- Nelson, E.C., Batalden, P.B., Plume, S.K., Mihevc, N.T. and Swartz, W.G. Report cards or instrument panels: who needs what?. *The Joint Commission Journal on Quality Improvement* 1995; **21**(4): 155-166.

NHS Executive *The New NHS - Modern, Dependable: A National Framework for Assessing Performance*. Consultation Document NHS Executive, January, 1998.

Nutley, S. and Smith, P.C. League tables for performance improvement in health care. *Journal of Health Services Research and Policy* 1998; **3**(1): 50-57.

Rosenthal, G.E. and Harper, D.L. Cleveland Health Quality Choice: a model of collaborative community-based outcomes assessment. *The Joint Commission Journal on Quality Improvement* 1994; **20**(8): 425-442.

Smith, P. Outcome-related performance indicators and organisational control in the public sector. *British Journal of Management* 1993; **4**(3): 135-151.

Smith, P. On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration* 1995; **18**(2/3): 277-310.

Stewart, J.D. The role of information in public accountability in A. Hopwood and C. Tomkins *Issues in public sector accounting*, Oxford: Phillip Allan, 1984.

UK Government *The Citizen's Charter: five years on*, London, HMSO, 1995.