BENCHMARKING AND INCENTIVES IN THE NHS

Paul A Grout, Andrew Jenkins
and Carol Propper

BENCHMARKING AND INCENTIVES IN THE NHS GROUT, JENKINS & PROPPER

BENCHMARKING AND INCENTIVES IN THE NHS

Paul A Grout, Andrew Jenkins
and Carol Propper

Centre for Market and Public Organisation
University of Bristol

Office of Health Economics
12 Whitehall London SW1A 2DY

ISBN 1 899040 41 2
Printed by BSC Print Ltd., London

Acknowledgements
We are grateful to Mary Bowerman, Diane Dawson, Matt Hinton, Neil Soderlund and Andrew Street for providing us with information about their research on benchmarking. Jon Sussex, Adrian Towse and three referees commented on earlier versions of this paper, suggesting many improvements. We accept full responsibility for the views expressed in the paper and for any errors that remain.

About CMPO and the authors
The Centre for Market and Public Organisation (CMPO) conducts academic research and contributes to policy debates on economic activities at the boundaries of the state. Founded in 1999, the Centre's core funding has been provided by the Leverhulme Trust, and it is based in the University of Bristol. Researchers at the Centre have published papers in areas such as health economics, regulation, poverty, and incentives in organisations. Further details are available from the Centre's web site at http://www.bris.ac.uk/Depts/CMPO/.

Paul Grout is Professor of Political Economy at the University of Bristol and Director of CMPO. His research interests include competition policy, utility regulation, public/private partnerships and industrial organisation.

Andrew Jenkins is a Research Assistant at CMPO. His research interest are in utility regulation and the economics of education.

Carol Propper is Professor of Economics of Public Policy at the University of Bristol and Deputy Director of CMPO. Her current research interests include the economics of health and health care and poverty dynamics. She was Senior Economic Adviser to the Department of Health on the Internal Market 1993-4. She is a co-director of the ESRC Centre for Social Exclusion at the LSE and a research fellow of the Centre for Economic Policy Research.

OFFICE OF HEALTH ECONOMICS

Terms of Reference

The Office of Health Economics (OHE) was founded in 1962. Its terms of reference are to:

- commission and undertake research on the economics of health and health care;
- collect and analyse health and health care data from the UK and other countries;
- disseminate the results of this work and stimulate discussion of them and their policy implications.

The OHE is supported by an annual grant from the Association of the British Pharmaceutical Industry and by sales of its publications, and welcomes financial support from other bodies interested in its work.

Independence

The research and editorial independence of the OHE is ensured by its Policy Board:

Peer Review

All OHE publications have been reviewed by members of its Editorial Board and, where appropriate, other clinical or technical experts independent of the authors. The current membership of the Editorial Board is as follows:

Professor Tony Culyer – *Department of Economics and Related Studies, University of York*
Professor Hugh Gravelle – *Centre for Health Economics, University of York*
Mr Geoffrey Hulme – *Director, Public Finance Foundation*
Professor Lord Maurice Peston – *Professor of Economics, Queen Mary and Westfield College*
Professor Carol Propper – *Department of Economics, University of Bristol*
Mr Nicholas Wells – *Head of European Outcomes Research, Pfizer Ltd*
Professor Peter Zweifel – *Socioeconomic Institute, University of Zurich*

CONTENTS

6

EXECUTIVE SUMMARY

This paper examines general issues in the use of benchmarking as a measure of comparative performance, reviews the application of benchmarking in the public and private sectors, and then examines the application of benchmarking in the UK National Health Service (NHS). A key issue of concern is whether there should be a link between benchmarks and explicit financial rewards, and the implications for the form of the benchmark if such links are made.

Our main conclusions are:

- There are strong arguments for the use of high-powered incentives attached to benchmarks in the NHS.
- In the light of experience in utilities we suggest that there is a danger of waiting for an ideal system. Introducing less aggressive high-powered incentive structures quickly is the best way forward.
- Further statistical modelling in the NHS is needed to isolate the impact on cost and quality of factors under the control of decision-makers. Compared to utilities there are many more data points in NHS activities so modelling should prove fruitful.
- The greatest difficulty for the use of high-powered incentives arises from the asymmetry in measuring the different tasks that organisations and individuals in the NHS are required to do. A combination of high-powered incentives and asymmetry of measurement can distort effort away from tasks that are harder to measure. For this reason, even in the very long run, incentives will almost inevitably be less high-powered for many NHS activities than elsewhere in the public and private sectors. To ignore this trade-off and push 'too hard' may have damaging side effects.
- The conveyance of information about objectives and mission is an inevitable consequence of a benchmarking regime. There needs to be rapid movement towards measures of quality adjusted for case mix (e.g. risk adjusted mortality outcomes). Failure to do this would convey the signal that the NHS is primarily interested in providing a cheap service rather than in providing a quality service at value for money.
- The number of benchmarks faced by any part of the NHS needs to be limited. The use of many benchmarks for each NHS

organisation creates the impression that the centre doesn't know what its mission is. This is at best confusing, but worse may lead to individuals reducing total effort. Successful agencies pursue narrow and clear missions. This suggests a limited number of benchmarks should be set for each type of organisation within the NHS, with different benchmarks set for different organisations.

- Systems take time to bed down so once in place should not be subject to continuous change.

1 INTRODUCTION

This paper examines general issues in the use of benchmarking as a measure of comparative performance, reviews the application of benchmarking in the public and private sectors, and then examines the application of benchmarking in the UK National Health Service (NHS). Benchmarking is a term used to cover a wide range of activities, which have in common the idea of comparison in order to identify opportunities for making efficiency gains. It can take a variety of different forms according to what level of activity is being compared, what kind of comparisons are made, and how the benchmarks are combined with allocation and reward mechanisms.

Benchmarking originated in the private sector, but has come to be used in the public sector as a means of improving performance for organisations thought to have weak incentives for efficiency. In a similar spirit, it is used as a means of promoting competition between regulated private sector monopolies (where it is also known as 'yardstick competition'). In the NHS, benchmarking is being developed rapidly at the same time as there is renewed interest in methods of incentivising decision makers within the NHS. One key issue is therefore whether there should be a link between benchmarks and explicit financial rewards, and the implications for the form of the benchmark if such links are made.

The paper focuses on the issues that arise when using benchmarking for this purpose. We identify the key factors which will be relevant in making an assessment of whether benchmarking is an appropriate tool for a sector, and if so, what form the benchmarks will take. In particular, we devote significant discussion to the analytical issues that determine the decision to link benchmarks to high-powered incentives, by which we mean material financial rewards attached to specific benchmarks. We then discuss how these issues affect the application of benchmarking in the NHS, again with a particular focus on the link between benchmarks and financial rewards.

Chapter 2 provides a brief introduction to the subject, identifies the key factors that are relevant in determining the applicability of benchmarks, and examines the issues identified by economic theory

that determine the form for the benchmark and the strength of incentives. Chapter 3 provides a wide-ranging review of the application to date of benchmarking in both the private and the public sectors. The experience in these sectors illustrates the similarities and key differences in the application of benchmarks in the private sector, as a tool for regulation of privatised utilities, and as a means of increasing efficiency in the public sector. We identify considerable similarity of issues that arise when benchmarks are used in public sector organisations that have multiple activities and many goals.

Chapter 4 examines these issues for the NHS. After a brief review of the progress in establishing benchmarking in the NHS, we focus our attention on three issues: the problem that benchmarking is trying to address; the form that the benchmark should take; and whether benchmarks should be linked to financial incentives, given the current institutional arrangements of the 'New NHS'. Our conclusions are that there is scope for the use of high-powered incentives attached to benchmarks in the NHS, and this will determine the form of the benchmarks to be used. But the existence of considerable asymmetry of information, and the fact that most if not all organisations within the NHS have multiple tasks mean we should be careful that such incentives are not pushed too far.

2 ANALYTICAL ISSUES

The purpose of this chapter is to provide a brief introduction to benchmarking, identifying some key factors that are relevant to an assessment of whether benchmarking is appropriate for a sector and if so, what form this should take. In particular, we discuss the issue as to how far to push high-powered incentives, (by which we mean material financial rewards attached directly to specific benchmarks), as part of the overall benchmarking strategy. We do not engage in the detailed specification of benchmarks. While clearly important, this is conditional on and hence secondary to the choice of overall framework.

2.1 What is benchmarking?

Benchmarking can be thought of as the comparison of business practices and performance levels between organisations in order to identify opportunities for making improvements. It can take a variety of different forms according to what type of activity is being compared, what kind of comparisons are made, and how the benchmarks are combined with allocation and reward mechanisms.

One common distinction is between metrics benchmarking and process benchmarking (PA Consulting Group, 1999). Metrics benchmarking, (also sometimes known as results benchmarking), can be defined as the quantitative measurement of inputs, outputs, outcomes and the relationships between them. Process benchmarking may use metrics as a starting point but goes on to discover the processes, systems, skills and technology which can be adopted in order to improve performance. Some commentators regard metrics benchmarking by itself as a rather limited exercise because, although it is useful in revealing the particular areas of weakness within an organisation, it does not explain the reason for weakness or suggest ways in which these areas could be improved. (PA Consulting Group, 1999; Holloway *et al*, 1998; Bullivant, 1996, 1998). On the other hand, it can be argued that one advantage of metrics benchmarking is that it provides information about the organisation as a whole. How efficient is the organ-

isation? How successful is it, compared to others operating in a similar environment? Process benchmarking looks at only one particular part of the organisation – the warehousing function or the in-house catering service, say. Concerns have also been expressed about the costs and time involved in conducting this form of benchmarking. There appear to be differences in preference for types of benchmarking as between the public sector, favouring metric benchmarking, and the private sector, where there is more process benchmarking, (see Chapter 3 below).

It is also common to distinguish the types of activity that are compared as follows:

(i) internal benchmarking – a comparison of similar processes within a particular organisation. The attractions of internal benchmarking are the ease of access to information and the low cost of a benchmarking project. However, by its nature it is unlikely to identify best practice.

(ii) competitive benchmarking – comparisons with an organisation's direct competitors. The most obvious form of competitor benchmarking is to look at one's genuine competitors. Although this provides information that is far more relevant than internal benchmarking, it is often difficult to obtain. One way of avoiding this problem is to benchmark against companies that are similar but not in direct competition. For example, electricity and water companies will often develop close relationships with companies in the same industry in a different country to avoid the immediate competition issue. Third parties may use benchmarks to create 'pseudo-competition' between organisations, particularly where there is weak formal competition, e.g. comparator competition in the water industry (see Section 3.2).

(iii) functional benchmarking – a comparison of specific business functions with practices and performance in organisations in other industries. The attraction of looking beyond industry boundaries is that one is more likely to identify best practice.

(iv) more generic benchmarking – external comparisons with businesses representing best-in-class for each particular aspect of the organisation's operations. The big disadvantage here is the cost and time involved.

Functional and competitive benchmarking are most common (Holloway *et al*, 1999).

2.2 What is the problem benchmarking is to address?

A frequently overlooked aspect of benchmarking is the articulation of exactly what the problem is that benchmarking is intended to address. In many respects this is the hardest part of the exercise. Hence the debate in the utilities as to how to use benchmarking, to create explicit competitive incentives where they are lacking, has received considerable attention.

The assessment of the core problem may have huge impact on the type of model that is implemented. We can identify two alternative problems that make this point very clearly.

Diffusion of good practice

The core problem could be that there is very slow diffusion of new ideas and best practice even though employees are well motivated to achieve the goals of the organisation. Strong motivation does not necessarily have to come from immediate financial rewards. Promotion within the organisation may be the prime motivator. In the public sector, through self-selection, employees have indicated that they have at least as high an interest in the area where they contribute as they do in monetary rewards. Health care workers, for example, are usually well motivated to supply quality healthcare. In such a position, if quality of practice is very different then the transfer of practice from the best to lower achieving groups may be a significant goal of benchmarking. Transmission of best practice may be a particular problem in the public sector since the job market tends to be less fluid, especially where the achievement level is below average. Best practice

within the private sector can more readily be transmitted through the movement of employees.

Benchmarking clubs (where groups of similar organisations get together to exchange information) and cross-fertilisation of ideas will work where employees are motivated, albeit in part by personal career concerns, to implement better strategies. Identifying and naming those using best practice will be beneficial since it enhances the incentive system, i.e. career concerns, and may encourage innovation by the best practitioners. In contrast, high-powered incentives, i.e. linking financial rewards to the benchmark, may be counter-productive since it can reduce the willingness of groups to share information or even dedicate time to other group sharing activities. High-powered financial reward systems may distract attention in such organisations from this sharing of good practice and towards short term cost cutting.

Poor incentives in the public sector

The core problem may alternatively be one of weak incentives. That is, career concerns are not strong enough to motivate most employees and there are insufficient incentives to deliver best practice. Creating the opportunities for diffusion of best practice will not be the solution in this case. Poor incentives are frequently portrayed as the core problem of the public sector arising from, amongst other things, lack of competitive pressures, absence of the profit motive, low salaries, and limited flexibility in the labour market. If this is the core problem then high-powered incentives may be appropriate since, putting to one side the practical problems of identifying the appropriate statistics to collect and use as a basis for reward, they replicate competitive pressures.

In practice, neither of these extremes is likely to paint the full picture. In some areas of the public sector individuals have very powerful career concerns and are well motivated but these concerns are different from those of their employer. An example arises in higher education where many academics are heavily driven by their research interests and far less by teaching but universities are motivated by stu-

dent numbers. Similar problems arise in health where hospital specialist doctors are driven by research and academic reputation whereas NHS Trusts are mainly interested in delivering patient care.

There is no reason to suppose that the same problems will apply throughout a sector or within an organisation. But for any given component it is important to understand whether diffusion of good practice or poor incentives is the bigger problem since the relative significance of these will dictate the approach to benchmarking that ought to be adopted. As we have indicated, pushing benchmarking clubs where incentives are low or individuals are focused on different issues will have little effect. Conversely, applying high-powered incentives where it is diffusion of best practice that is needed may deflect from this activity. Articulation of the core problem in specified areas may not be easy but it is an essential precursor to the design of a sensible and focused benchmarking regime.

2.3 The role of high-powered incentives

A central question is how and to what extent should financial rewards be linked to the benchmark? Although there is a large and ever growing literature on incentive schemes there are no simple answers to this question. Indeed, in the context of the public sector in general, and health in particular, there are good reasons for thinking that this problem is difficult to answer. It is useful to consider some of the abstract ideas that are relevant to this issue.

The principal-agent problem

At an abstract level it is useful to think of most of those working in an organisation as the 'agents' of a 'principal' where the latter is responsible for setting the goals of the organisation. The principal wishes to reward the agent if it achieves the principal's goals, e.g. lower cost, greater output, rather than if the agent follows its own objectives, e.g. lower effort, higher discretionary expenditure. It is essential, however, when designing a scheme, to take account of the fact that the agent

normally has more information than the principal. Consider the following example. When a principal designs an incentive scheme it has to accommodate the fact that in some circumstances costs will be higher than average because effort (for example to control costs) is lower than average. Hence the principal wants to penalise agents that have high costs. In some circumstances, however, costs will be higher than average for factors that are outside the agent's control. In these cases penalising high cost will lead to a worse service (as effort is put into recovering the penalty by cutting services) and demoralise staff. This creates a conflict for the principal. Aggressive schemes that reward above average performance strongly and penalise below average performance will provide strong incentives to those that can increase effort but do damage to those agents that have limited scope to improve things. On the other hand schemes that attach no reward for above average performance will avoid damaging those who have high costs for reasons outside their control but provide no incentive to the others to improve performance. A balance has to be struck. To provide reasonable incentives for staff in the latter case it is essential that the incentive regime should not be too aggressive on higher than average cost performers. That is, it is not possible to design as effective a focused incentive regime as would be possible if one could be sure that higher than average cost could be attributed solely to low effort.

The core problem as described here provides a series of insights. First, it is essential to recognise when designing schemes the limitations on what can be achieved rather than have too high expectations of the role of high-powered schemes. Indeed, failing to recognise the problem and instead implementing schemes that are too high powered for the situation can have detrimental rather than beneficial results. Second, a scheme that provides a balance between the conflicting objectives inevitably creates a rent, or surplus, for those that are able to reduce relative costs by increasing their effort. This is sometimes called an 'information rent'. It follows that one cannot bring about significant improvements in effort in a costless way. Real resources need to be committed and it has to be accepted that those agents who

achieve below average costs will have to earn an abnormal return. That is, a good scheme must involve sharing of rents between principal and agent. Third, it is important to understand the extent to which there is asymmetric information in the system, with agents having access to more information about their own costs than principals do. If there is very little, i.e. it is easy to identify many of the innate differences in relative costs, then it will be possible to implement quite high-powered incentive schemes that should be very effective at low cost. In contrast, if there is an enormous amount of uncertainty, i.e. the principals find it very difficult to understand accurately what is causing the differences in costs and whether these differences are in the agent's control, then the link between financial rewards and relative performance will have to be muted. We can expect the effects on effort to be muted as well as a consequence.

The role of benchmarks

A critical variable in all of this is the quality of data. It is perfectly feasible to introduce financial incentive schemes that pay no regard to comparative data. The big advantage of using benchmarking as the reward mechanism is that 'feedback' is reduced. Feedback arises when the actions of the agent lead to changes in the regime that take away the agent's reward. For example, if cost savings feedback quickly into lower budgets or harder targets then incentives to reduce costs are weak. On the other hand, the principal would, at least on average, like to capture benefits quickly. Using performance against other organisations it is possible to eliminate both feedback and, to some degree, lack of information. These are in essence two sides of the same coin, i.e. using external information by definition eliminates the feedback problem. A difficulty with incentive schemes that do not use comparative data is that a benchmark, in its broader sense, has to be set for acceptable performance. Inevitably, the initial position of the organisation will have a strong effect on the incentives. There are two particular reasons for this. One is that excellent performance against a benchmark in the previous period is more likely to make good perfor-

mance difficult to achieve in the current time period since the organisation is likely to be close to optimum efficiency already. Second, good performance affects the principal's view of what can be achieved so that the current performance target may be raised in the light of past performance. This 'ratchet effect' can be a significant problem and means that the agent may be better off playing 'games' against the principal by holding back potential gains and releasing them slowly.

Setting financial incentives as a function of relative performance avoids some of the problems mentioned above. For example, excellent performance in one period will only shift the aggregate by the weight of the agent in the overall benchmark. The feedback effect will be small if there are many agents in the benchmark and so incentives are preserved. Similarly, if there is no feedback onto targets from an agent's effort then all have stronger incentives to reduce costs, or meet whatever is the target benchmark, so those that perform badly are penalised as the benchmark moves ahead. This type of direct financial reward according to relative performance is called yardstick competition or comparator competition (Shleifer, 1985)[1].

How successful yardstick competition can be depends on the extent of the information problems discussed above. If there are agent specific differences that cannot be influenced by the agent's effort then these either should be accounted for in the individual benchmark or the reward system needs to be less stark to provide some protection for the disadvantaged agents. A benefit of using comparative performance is that some of these specific differences can be picked up in the overall benchmark through statistical modelling of the cost conditions. By definition, however, if one is using an agent specific incentive scheme, often called a bottom up approach, then there can be no scope for this. Of course, truly agent specific differences can in part be picked up by

1 The US DRG payment system is very similar to yardstick competition in that it provides the same payment for the same treatment regardless of which hospital the treatment takes place in. In principle, at least, this will reward those hospitals which are relatively efficient since they will reap a larger surplus than inefficient hospitals.

statistical techniques if one has data on each organisation over several years. Where this is lacking, or not of sufficient quality, 'judgement' has to be used. This involves application of the principal's views to deal with special aspects that cannot be addressed objectively. The water regulator for England, for example, is quite explicit that judgement is an important part of comparator competition. The judgement process may reintroduce the feedback problem through the back door. This is an issue that the UK utilities and their regulators have grappled with since privatisation (see Section 3.2).

A clear distinction arises in the use of benchmarks between those where a natural comparator exists and those where this has to be constructed. Where there is a simple comparator, e.g., profit, sales, cost subject to a simple quality standard, then benchmarking with high-powered incentives can work well. Where life is more complicated the difficulty in identifying a sensible benchmark is often a major problem particularly if rankings of organisations are sensitive to the comparator that is constructed. The problem of identifying appropriate benchmarks has been faced in the regulation of the water and electricity industries, where benchmarking is used to help set price controls. It has been found that the relative performance of companies is sensitive to the exact benchmark used. In addition, identifying an output measure to assess relative costs of companies has been extremely difficult in part because of the small number of companies and observations. This is discussed in Section 3.2.

We have indicated that the use of comparative data may reduce some of the gaming activity such as deliberate under-performance to prevent tough targets in future rounds. However, in general gaming is more likely to be increased by the introduction of benchmarking targets particularly where there is poor information. Reward follows good performance as measured by the benchmark and if, following introduction of a scheme, it is now worthwhile putting in greater effort to improve true performance it is inevitable that it will also be worthwhile putting in effort to present the evidence in ways that raise measured performance even when there is no true improvement. An

obvious example has been the response to NHS waiting list targets. Some of the decrease in waiting lists and times has been due to increased levels of activity and some due to manipulation of the date at which individuals are recorded as being put on the list. One needs to be confident one is measuring something worthwhile but gaming is an inevitable consequence of high-powered incentives. Whilst schemes should try as much as possible to limit this and to choose benchmarks that are less susceptible to manipulation, no benchmark can avoid this problem. Introducing more and more restrictions in the face of evidence of gaming by agents may make the final scheme too complex and restrictive to achieve the objective.

Multi-tasking

An agent is unlikely to be engaged in one single activity. Where the output of the agent is varied then there are many outputs and inputs that the principal may wish to benchmark and reward. A situation where an agent is involved in various activities is referred to as multi-tasking (see Williamson, 1985; Holmstrom, 1979; Holmstrom and Milgrom, 1991; and Tirole, 1994). A problem arises if some of the tasks are harder to measure than others. If all are easy to measure then a composite benchmark can be used, albeit after resolving the problems we have considered above. However, while theory may imply that that the problem is easy to solve if all tasks are measurable, in practice agencies may simply lose focus and direction if there are too many tasks to be measured and rewarded. It is probably better to avoid too many targets.

The real problem with multi-tasking arises if some of the tasks are harder to measure than others. In these circumstances a high-powered incentive scheme that is successful in encouraging greater effort is not unambiguously desirable. The reason is that part of the better performance may be the result of diverting effort and attention away from those tasks that are hard to measure and which are not part of the high-powered incentive scheme. To prevent a reduction in effort expended on tasks that are hard to measure, a sensible strategy is to be

less aggressive with incentives. That is, the presence of multi-tasking attenuates even further the role of financial rewards.

A disputed example arises in education. Schools help children to pass exams. The success of schools in doing this can be measured very easily. Schools also provide a means of socialising children: getting on with other people, making moral judgements, and being good citizens. This is very difficult, perhaps impossible to measure. But only using exam results can have various perverse consequences: neglect of other parts of good schooling, neglect of children not likely to pass exams or of high academic ability who will pass easily, and dis-enrolling less academic pupils in order to distort the measure (see Section 3.5 below). This is a clear multi-tasking issue. The extent to which it is a problem depends on the weight that is placed on academic achievement versus the other, more general, outputs.

Similar behaviour may have arisen as a result of NHS waiting list targets. To meet such measured targets, Trusts have concentrated their efforts on reducing these lists, possibly at the expense of other, more difficult to measure, health care activities.

The problem of multiple principals

So far in this subsection we have discussed principal-agent problems. However, it is not obvious that all problems with benchmarking fall into this simple categorisation. There is no reason to suppose that a particular group will only have one principal. An obvious example arises if an employee receives income from more than one source. More subtle examples are probably more common. Although it is standard that there may only be one employer that provides a salary, employees may feel that significant drivers in their decisions and controls on their careers come from other sources e.g. professional bodies. Multiple principal issues arise in the health care sector, for example, because clinicians may owe loyalty to both the NHS Trust which employs them and to a professional body such as a royal college or the British Medical Association. This problem of multiple principals brings with it further difficulties for providing optimal incentives (see

Dixit, 1997). If each principal sets its reward system to further its own objectives, then this will be achieved by providing incentives to follow its scheme and so offset the financial rewards of other principals. The net effect may tend to cancel out the incentive schemes. Indeed, the agent can end up with completely minimised incentives because the schemes offered by principals work against each other in a way that the agent is protected by the array of incentives against downside risk that would otherwise arise from low effort. A solution, if it can be achieved, is to further reduce the extent of high-powered incentives.

Group versus individual rewards

There are several arguments for team rewards. Teams may be a more efficient way to deliver a product, as they may encourage co-operation and inter-professional working, and they allow monitoring by groups of employees who can have better information than managers, so discouraging shirking and increasing effort. An individual reward scheme placed in such a context could be arbitrary, costly to operate, and possibly encourage responses that lowered or diverted effort from, for example, the production of health care. On the other hand, simple economic theory suggests that team rewards are prone to free-riding, and that the free-riding problem becomes worse the larger the team.[2] However, the fact that many establishments do operate group performance related pay (PRP) schemes, however, suggests that they do provide a significant incentive for employees (see e.g. Drago and Heywood, 1995). Kandel and Lazear (1992) investigate some alternative ways in which this problem may be resolved in team production settings by means of peer pressure. Peer pressure translates into incentives by punishing workers who deviate from what is expected of them by guilt and/or shame. When team workers are able to monitor each other's efforts more easily than can a third party, the firm may offer a group PRP contract which induces team members to apply peer pres-

2 For a recent discussion, see Auriol *et al* (1999).

sure or which induces feelings of guilt when workers put in too little effort. Kandel and Lazear hypothesise that profit sharing creates empathy towards those who receive the residual profit. Workers more readily empathise with other workers than with faceless shareholders. Further, the more empathy there is towards the joint beneficiaries of one's effort, the greater is worker motivation.

The issue of group versus individual rewards is clearly an important one in health care, where the output is very much a team effort and where the importance of teams has come to be stressed more in recent years.

2.4 Benchmarking as a statement of mission

Benchmarking is rarely thought of as conveying information out from the centre. It is thought of as conveying comparative information to the centre. But the ability of benchmarks to convey information in the opposite direction can be very effective. Kaplan and Norton (1996) identify a growing focus in design of measurement systems on how clearly the system communicates the organisation's strategy (see Eccles, 1991; Burns, 1992; Locke and Latham, 1996; and Courty, 1997). Kaplan and Norton point out that a coherence between the objectives of a company and what is measured is crucial: 'Those companies that can translate their strategy into their measurement systems are far better able to execute their strategy because they can communicate their objectives and their targets'.

An important point not explicitly made in this literature is that conveyance of information about objectives and mission is an inevitable consequence of a benchmarking regime whether it is an explicit objective or not. The decision to measure specific outputs or inputs and not others conveys messages to those in the organisation about what the centre thinks is important. Even the decision to introduce benchmarking is itself a message about the perceived mission and thinking of the centre. The identification of best practice and encouragement of groups to transmit this information can provide a strong

signal that the centre's mission is to promote excellence. In contrast, a strong focus on cost reduction without any measurement of output quality indicates to the organisation that the mission puts greater weight on cost reduction than quality.

If the benchmark is changed frequently then there can be no clarity of message from the centre. At best this can lead to confusion but it can also be extremely de-motivating. Clearly, introduction and frequent changes of benchmarks without adequate thought to the signal this transmits can be damaging but going too slowly and not implementing appropriate benchmarks can display a lack of commitment at the centre. The active collection of performance indicators is a sign that the organisation is devoting resources to checking progress, and is a clear indication that these are seen as being at the heart of the mission.

It is also worth noting that benchmarking, in the sense of cross-sectional performance data, can contribute to the acceptance of the corporate vision in addition to its role as a signal of that mission. In the absence of data, individuals within the firm may believe that the organisation is performing better than it is. There is some evidence, from both the public and private sectors, that managers tend to overestimate the strengths of their organisation (Voss *et al*, 1997; Bovaird and Davis, 1999, p 308). Benchmarking data can challenge this kind of corporate complacency by providing hard evidence that the organisation is under-performing relative to competitors or best practice. This will motivate managers in the organisation by providing them with a truer picture of what their organisation is really like, and the size of the gap to be closed in order to attain best practice performance levels. Benchmarking can therefore contribute significantly to 'organisational learning' (Voss *et al*, 1997).

2.5 Summary

We have identified a series of factors that affect the choice of benchmark and its use but there are three that we have identified that we

believe are of particular importance:

- It is essential to analyse carefully what the problem is that benchmarking is designed to solve. This stage has a critical effect on the type of benchmarking that should be put in place.
- High-powered incentives are important in the right circumstances. Where it is clear that low effort is the reason for poor performance and where it is easy to identify poor performance that is the result of factors outside the agent's control, then high-powered incentives are appropriate. If it is harder to distinguish the reasons for poor performance then the incentive regime should not be too hard on poorer than average performers. The more the uncertainty, the less aggressive the scheme should be. Similarly if an agent engages in multi-tasking where some tasks are hard to measure then the incentive scheme rewarding better than average performance on the measurable tasks should be less aggressive to limit the agent's desire to switch effort away from non-measured tasks.
- The conveyance of information about objectives and mission is an inevitable consequence of a benchmarking regime whether it is an explicit objective or not. It is essential to take this into account when determining benchmarking strategy.

3 SECTORAL EXPERIENCE OUTSIDE THE NHS

In this chapter, we describe and evaluate benchmarking in a range of sectors. We begin with the private sector (outside utilities), where benchmarking was first introduced, then look at the regulated utility sector, where a system of comparative competition has been in use for several years, and then examine in turn the experience of benchmarking in central government, local government and in education. In each case we examine the extent and experience of benchmarking, focusing on the question of who instigates the benchmarking, the problem that benchmarking is being used to solve, the form and level of sophistication of benchmarking, and the incentives for organisations in the sector to adopt benchmarking. Our objective is to build up a set of practical alternative approaches to benchmarking which can inform the debate on how best to conduct benchmarking in the health sector.

3.1 Private sector

Experience of benchmarking

Although for many years organisations have sometimes used informal comparisons to improve their own performance, the formalisation of comparative analysis into a set of clearly-defined steps, known as benchmarking, is usually traced to Rank Xerox. In the late 1970s and early 1980s Xerox realised that it was being out-competed. The retail price of Canon photocopiers was less than Xerox's manufacturing cost. Increasing competition from Japanese firms reduced Xerox's market share from 67 per cent of the plain paper copier market in 1976 to under 45 per cent in 1982.

Rank Xerox set out to learn good practice from the Japanese. A team was sent to Japan to compare Xerox's performance in a wide range of areas with their competitors, including Fuji-Xerox, their part-owned subsidiary in Japan (Dence, 1995; Coopers & Lybrand, 1994). Measurements were taken of such things as production costs, cycle time, and overhead costs and the performance of the company was compared to its main competitors. As a result, Xerox was able to improve quality and cut costs.

Table 1 **The Xerox 10-step approach to benchmarking**

Planning
1. Identify benchmark outputs
2. Identify best competitor
3. Determine data collection method
Analysis
4. Determine current competitive gap
5. Project future competitive performance levels
Integration
6. Communication of data; acceptance of analysis
7. Develop new goals and functional action plans
Action
8. Implement specific actions
9. Monitor results and report progress
10. Re-calibrate benchmarks

Source: Dence, 1995.

The Xerox approach to benchmarking was subsequently codified into a 10-step procedure. This approach is set out in Table 1. It, and variations on it, have been widely adopted.

During the 1980s and 1990s, benchmarking spread rapidly. The success of Xerox and other companies which used benchmarking to improve performance attracted media attention. Benchmarking became the latest management fad, while consultants enthusiastically promoted it. The publicity surrounding prestigious quality awards, in which benchmarking plays a significant role, may also have helped to increase awareness of benchmarking amongst business people (Whymark, 1998).

A survey conducted for the Confederation of British Industry (CBI) in 1992 found that 67 per cent of companies claimed to be doing benchmarking (Coopers & Lybrand, 1994); this had risen to 78 per cent in 1994 and 85 per cent in 1996 (results cited in Davies and

Kochhar, 1999; Ahmed and Rafiq, 1998). However, these surveys were undertaken on a sample of the largest 1,000 UK companies. Such companies are most likely to be engaged in benchmarking activity. A survey by the Open University Business School (OUBS) which was conducted in 1997 found that only 12 per cent of firms with fewer than 25 employees had undertaken benchmarking activity; this rose to 23 per cent of firms with 26-99 employees (Holloway *et al*, 1999). Other surveys have also found that less than 20 per cent of small and medium sized firms have engaged in benchmarking activity (results cited in Holloway *et al*, 1999, p 18; see also Monkhouse, 1995).

Although benchmarking originated in the private sector, during the 1990s the development of league tables in education and of Audit Commission comparative performance indicators has meant that benchmarking is now more common in the public sector than many parts of the private sector (as shown in Table 2). The data in Table 2 suggest that private sector benchmarking is most widespread in the utilities, where comparative analysis of costs is often imposed by regu-

Table 2 **Benchmarking activity by sector in 1997**

Sector	Total number	Number claiming to be benchmarking	Per cent
Government	55	32	58
Education	37	23	62
Health	52	36	69
Manufacturing and construction	269	135	50
Financial services	57	19	33
Services and retailing	189	68	36
Utilities	18	14	78
Other	49	19	39
Missing data	7	n/a	n/a
Totals	**733**	**346**	**47**

Source: Holloway *et al*, 1999, p 21.

lators (as discussed in Section 3.2), and that it is more common in manufacturing than in services, probably because output is easier to measure in manufacturing.

What is the problem that benchmarking is trying to address?
Private sector companies use benchmarking with the aim of improving their competitive performance. Benchmarking can contribute to this goal in several ways. Metrics benchmarking will provide valuable information to the firm about how well it is doing relative to its competitors in particular markets. Do other companies have lower costs, or higher levels of customer satisfaction? An example of the kind of data obtained from a benchmarking exercise is shown in Table 3, which refers to a successful functional benchmarking initiative by Rank Xerox of its logistics and distribution (L & D) department beginning in the early 1980s (Tucker *et al*, 1987). In the past, this department had achieved labour productivity increases of 3-5 per cent per annum. By the 1980s, industry price cuts implied that further improvement was necessary to maintain profit margins. After searching trade journals, and discussions with professional associations and consultants, LL Bean Inc, a sporting goods retailer and mail order company, was selected. Comparisons of Xerox and LL Bean distribution centres, both of which had systems designed to cope with products diverse in size, shape and weight, revealed weaknesses in the Xerox warehouse. The comparisons below show performance at Xerox's most efficient warehouse compared to LL Bean in 1982.

Table 3 **Comparison of key performance criteria in two distribution centres**

	LL Bean	Xerox
Orders per person day	550	117
Lines per person day	1,440	497

Source: Tucker *et al*, 1987.

A line represents picker travel distance for one trip to a bin, and is a key measure of productivity.

Comparative data from metrics benchmarking also has a useful role in setting targets. Chaparral, a US steel company which began in 1975 as a mini-mill producing simple reinforcing bars, began regular visits to steelmakers in Europe and Japan, to obtain data on measures such as equipment cycle times, yields and inventory levels. The processes underlying these metrics were also examined so that improvements could be implemented back in the company's US plants. The main use of the benchmarking data was to set targets for particular aspects of Chaparral's business (Dence, 1995). In the UK some 90 per cent of the large firms in the CBI survey reported that benchmarking enabled them to set meaningful and realistic targets (Coopers & Lybrand, 1994). Benchmarking data can also motivate staff in the company by showing them what is being achieved in other companies, and the data may also help to persuade senior managers in the company of the need to make changes.

Once benchmarking moves beyond metrics to comparisons of processes, it can provide a source of information on new ideas and alternative ways of carrying out tasks. In the Xerox/LL Bean example, personnel from the Xerox L & D department visited LL Bean to examine the processes which underlay differences in metrics. Important factors which helped to explain LL Bean's high productivity included arranging materials by velocity so that fast moving items were stocked closest to the picking route; sorting and releasing incoming items throughout the day to minimise picker travel distance; basing incentive bonuses on picking productivity offset by error rates. Some of the processes in use at LL Bean were subsequently introduced into the Xerox L & D department. (Tucker *et al*, 1987).

Benchmarking, then, has the potential to provide both quantitative and qualitative information to the firm in its search for best practice and competitive advantage. A survey by Drew (1997) of 140 North American firms reported on some of the benefits that benchmarking can provide. The sample was drawn from a broad range of sectors

Table 4 **Areas of success in benchmarking, mean responses**

Identifying creative and useful new ideas	5.50
Setting 'stretch' goals for improvement	5.48
Identifying best practices in industry	5.38
Improving customer service/quality	5.26
Convincing people of the need for change	5.23
Supporting business process redesign	5.09
Developing new products/services	4.29

Source: Drew, (1997). Sample size 140.

including manufacturing, financial services, high technology industries, services, government and healthcare. Respondents were asked to measure the benefits on a 1 to 7 scale, with 7 representing greatest value.

These results confirm that obtaining new ideas, setting clear targets for improvement, and identifying best practice are among the most important benefits of benchmarking.

Who is the instigator?

In the private sector individual firms instigate benchmarking activity. This contrasts with the public sector, where benchmarking is often imposed from above. This distinction between the public and private sectors is not entirely clear-cut: some public sector organisations do engage in benchmarking activity voluntarily, and in the utilities a regulator imposes benchmarking onto private sector firms. Nonetheless, it remains the case that the predominant pattern in the private sector is for firms to instigate benchmarking themselves.

Some firms do not instigate benchmarking. Small firms seem particularly unlikely to engage in benchmarking, often because they have not heard of it (Holloway *et al*, 1999; Partnership Sourcing, 1997). The Department of Trade and Industry (DTI) has established the UK Benchmarking Index which provides small businesses with a computer-based system for undertaking benchmarking. The small business fills in a questionnaire, which the business adviser submits to a com-

puterised database. This generates a report on the strengths and weaknesses of the small business which can be used as a basis for reviewing the firm's performance.

The form of the benchmark

In the UK most companies appear to have moved beyond purely internal comparisons. Survey data show more than 85 per cent of firms either using external, or both internal and external, benchmarking (Holloway *et al*, 1999). Comparison against direct competitors and others in the same industry are now the most common types of benchmarking undertaken by companies (Partnership Sourcing, 1997, p 12). However, very few companies were utilising the most sophisticated forms of benchmarking. Only eight per cent of firms in the OUBS survey were looking outside their own industries for benchmarking partners, and these tended to be very large organisations with a long history of benchmarking activity (Holloway *et al*, 1999, p 22).

In practice, many companies have found some difficulty with benchmarking. There were problems finding appropriate benchmarking partners, obtaining comparable information, and processing data when it was available. Companies also reported significant resource constraints, especially the time needed for benchmarking, difficulty gaining access to competitor organisations, and staff resistance to benchmarking was also encountered (Holloway, *et al*, 1998). Even among the large organisations in the CBI survey, over two-thirds of companies reported problems with both gaining access to confidential competitor information and the comparability of data from different companies (Coopers & Lybrand, 1994). Companies appear to make surprisingly little use of benchmarking networks and expert consultants (Holloway *et al*, 1998 p S123).

The effects of benchmarking

Anecdotal evidence sometimes suggests that benchmarking can lead to miraculous improvements. For example, a benchmarking consultancy reported that there were labour productivity differences of 50 per

cent between different building materials manufacturers, a 200 per cent productivity differential on some processes in different plants of a commercial foods company (cited in Dence, 1995). However, it is very doubtful that benchmarking always uncovers such large potential for improvement. There is a limited amount of more realistic evidence on the benefits of benchmarking available.

The best approach is to examine statistically whether firms which have done benchmarking perform better than those which have not. There appears to be only one study of this kind. Voss *et al* (1997) report the findings of a London Business School survey of 660 managers in the UK, Germany, the Netherlands and Finland. This survey was conducted in 1993/94 and looked at the manufacturing sector.

Respondents were asked to indicate the extent to which the manufacturing site used benchmarking, on a 1 to 5 scale, with 1 representing no benchmarking, and 5 representing 'regular, documented benchmarks against competition and world class standards beyond company's industry'.

To test whether there was a link between benchmarking and performance two indexes of performance were constructed, one of operational performance and one of business performance. The operational performance index was constructed by summing 17 items from the questionnaire covering quality, productivity and cycle time. Business performance was constructed from six items including high levels of customer satisfaction, an increasing market share, positive cash flow, return on assets better than that of competitors, low product costs, high productivity growth.

Regression analyses were performed to test the hypotheses that operational performance is positively related to benchmarking, and that business performance is positively related to benchmarking. They found that the extent of benchmarking was significantly related to both operational performance and business performance. However, no other variables which might explain performance were included. Also, as the authors admit, whether benchmarking leads to better performance or whether better performing companies use benchmarking

more, remains an unanswered question. Regression analysis reveals nothing about causality.

More common in the literature is to ask firms if their benchmarking activity has been successful. The Coopers & Lybrand survey reported that, of those engaged in benchmarking activity, 82 per cent regarded their benchmarking projects as successful; 14 per cent reported that it was too early to assess; five per cent reported 'don't know'. Perhaps not surprisingly, no benchmarking projects were reported as being unsuccessful. Furthermore, more than two-thirds (68 per cent) stated that they expected to increase their use of benchmarking over the ensuing five years, while 31 per cent expected no change, and one per cent was unsure (Coopers & Lybrand, 1994).

The OUBS survey asked a sample of 97 firms whether benchmarking had delivered the improvements in performance that they had anticipated. Responses were placed in categories from 1, 'not at all' to 7, 'exceeded expectations'. On this basis, the median response was 4 and the mean about 3.8, suggesting that firms tended to regard their benchmarking activity as reasonably successful, but not a massive success (Holloway *et al*, 1999).

Conclusion

There is therefore limited evidence evaluating what has been achieved by benchmarking. The literature has tended to be practitioner-oriented, with a range of meanings attached to the term benchmarking, making systematic analysis of the phenomenon difficult.

Nonetheless, benchmarking has now been in existence for 20 years or so in the private sector. It is no longer the latest fashionable management concept, but has become widely used and accepted as a business tool, especially by larger firms.

3.2 Utilities

Experience to date

The presence of sector specific regulators sets the utilities apart from

the rest of the private sector. They have used a form of benchmarking based on comparative cost data from the firms in their industries to set prices. This approach has been most fully developed in the water industry, where it is referred to as comparator competition, but has also been used in the electricity transmission and distribution industry. Some of the early privatisations, such as gas and telecoms, were privatised almost as single entities, which ruled out opportunities for regulators to undertake within-industry comparisons.

A relatively crude comparative cost analysis underpinned the initial price limits set for the water industry at the time of privatisation in 1989. Ofwat, the regulator, conducted a large amount of data gathering and analysis to improve the system prior to the first price review in 1994, and the system was refined further for the next price review which was completed in 1999. In addition, Ofwat has conducted international comparisons notably with Australian water companies (Ofwat, 1998a; Ofwat, 1999a). Although there is a good deal of data in the reports, the work is still at the experimental stage, as there are serious problems of comparability to be overcome. Successive reviews of the electricity distribution companies have also made some use of comparative cost data, but this has received more emphasis in the most recent, 1998/99, price review.

Independently, the companies have undertaken some benchmarking activity of their own. For example, in the electricity industry, a consultancy firm is used to exchange information amongst companies whilst maintaining data confidentiality. Companies use benchmarking in various sections of their business, e.g., Northern Ireland Electricity uses comparator companies in the electricity industry to benchmark customer service standards and network performance (MMC, 1997). Most water companies have used benchmarking of business functions such as customer services and metering. Ofwat is encouraging companies' use of benchmarking and recently commissioned a report by a group of consultants (PA Consulting Group, 1999). The main conclusion of the report was that the use of process benchmarking varied a good deal within the water industry and that there was plenty of

scope for improving customer service and reducing costs through greater use of process benchmarking.

What is the problem benchmarking is trying to address?

For the companies, benchmarking addresses the standard problem of seeking information on best practice to identify cost-cutting and profit-making opportunities. They see benchmarking in the same way that the rest of the private sector does.

For the regulators, comparative cost data are used to set price limits for the companies. Whilst some utility markets are becoming much more competitive there remain markets with monopoly power where benchmarking can be used to create pseudo-competition. In the water industry, for example, competition between suppliers will be very limited for years to come.

Under the system of regulation which has developed in the UK since privatisation, the regulatory body sets a ceiling on prices according to an RPI ± X formula, based on what it believes to be a reasonable rate of return. Usually, the regulator sets the price ceiling for a period of four or five years. If a company makes greater than expected efficiency savings it will be able to keep the additional profits generated during the period over which the price cap is set. Conversely, a company which is less efficient will earn a lower rate of return. Price cap regulation, then, provides incentives for companies to be efficient. The general idea behind this is of replicating competitive pressures in markets where real product market competition may be limited. Comparative competition aims to mimic the effects of a competitive market in which firms face commercial pressures to improve service and quality at prices that customers are prepared to pay. In markets, such as water, where genuine product market competition between rival suppliers may not exist, a system which mimics competitive pressures has the potential to improve performance.

However, unadjusted cost data cannot be used because the business environment varies from company to company. In order for regulators to fairly assess the efficiency of firms it is essential that factors

which influence results but are exogenous to management, such as geographical conditions, population density, regional variations in per capita income and so on, are allowed for in the analysis. There are a number of techniques available, which can be used to accomplish this.

Regression analysis is probably the most widely known technique. Here the sign and size of the residuals from a regression equation with cost as the dependent variable will reveal which organisations are of above average efficiency, and those of below average efficiency. An alternative is to use stochastic frontier estimation, which recognises the distinction between inefficiency and measurement error. The problem with this technique is that highly specific assumptions have to be made about the distributions of the composed error between inefficiency and measurement error. In practice, many studies find that the measurement error swamps inefficiency so that measurement error is found to be large and inefficiency implausibly negligible (Drake and Weyman-Jones, 1996). A simpler alternative is data envelopment analysis (DEA), a linear programming approach which fits a curve or envelope around outlying, efficient organisations. Efficient organisations will lie on the envelope curve, which defines the efficiency frontier, while less efficient organisations, will be some distance from the envelope curve and their distance from the envelope can be used to measure the degree of inefficiency.

Advocates of DEA, such as Smith (1990), argue that the technique is well adapted to handling multiple inputs and outputs, which cannot be readily accomplished by regression analysis. On the other hand, those undertaking regression analysis, unlike DEA, have the advantage of being able to use a number of well-developed statistical tests, to assess both the significance of explanatory variables and of the functional form. DEA is also sensitive to outliers, and may erroneously conclude that they are efficient (Cubbin and Tzanidakis, 1998). Although DEA and stochastic frontier analysis allow a ranking of firms according to their position relative to the frontier, in practice, regulatory bodies have tended to use mainly regression analysis.

The form of the benchmark and the role of incentives

Prior to privatisation, there had been no standard measure of output in the water industry, and the water companies' accounting practices were far from uniform. In the early years of the new regulatory system, a good deal of time was spent in sorting out standardisation of output measures and accounting information. Much of this was done in working groups on comparative efficiency comprised of personnel from the industry and from Ofwat (Sawkins, 1995). Ofwat issued several regulatory accounting guidelines to standardise the cost information it received from the water companies.

As a result of the consultations between Ofwat and industry experts in the early 1990s, a range of explanatory factors expected to have an influence on costs were identified including resource characteristics, population/economic growth, topography, and asset condition. By 1992 Ofwat was able to conduct a preliminary econometric analysis of the factors which affected companies' operating costs.

Improvements were subsequently made to the data, and Ofwat commissioned Professor Mark Stewart of Warwick University to conduct an econometric analysis of comparative efficiency for the 1994 Periodic Review. Again, this involved regressing operating cost expenditure on a number of explanatory variables. In assessing the efficiency of companies' operating costs for the Periodic Review in 1994, the Director General drew on the research of Professor Stewart updating it to take into account more recent data and comments from the companies. Ofwat's analysis indicated that, in the case of water service, the most important drivers of operating cost per unit of water delivered were the length of the distribution system, the amount of pumping needed, the proportion of demand from large customers, the level of treatment provided, and the size of the water treatment works.

The rankings derived with these procedures are grouped. It is the differences between the groups that are used for regulatory purposes. For example, the companies were grouped together into broad efficiency bands for the purposes of the 1994 Periodic Review where com-

panies were classified as either 'more efficient', 'around average' or 'less efficient' (Ofwat, 1994).

Ofwat has continued to adhere to the same approach since 1994, making some amendments and refinements in preparing for the most recent, 1999 review (Ofwat, 1998b). Operating expenditure and capital expenditure have continued to be treated separately, but capital expenditure has also been analysed econometrically, rather than by the standard costing approach used previously. The companies have been placed in one of five bands – A to E – for their operating efficiency and for their capital maintenance expenditure (Ofwat, 1999b).

The electricity regulator Offer (or Ofgem as it became known following merger with the gas industry regulator Ofgas) has not traditionally devoted much effort to comparative competition. However, in its current price review of the electicity distributors it regulates Offer has placed more emphasis on cross-company comparisons (Offer, 1999). The preliminary stage of this exercise has consisted in making the operating cost data as comparable as possible by adjusting for differences in accounting policies, cost allocations and attributions, regional factors and one-off costs. An important factor in determining electricity distribution costs is the pattern of peak demands at different points within the system of each electricity distribution company. These peaks cannot easily be measured, however, so a composite proxy variable for peak demands has been constructed. Offer has regressed base operating costs on the composite variable to assess the relative efficiencies of the companies. Offer regards the results as reasonably robust, but stresses that undue reliance should not be placed on the statistical analysis. In fact, it forms only part of its assessment of operating costs, which has been principally informed by the consultants' study which compared practices between companies (Offer, 1999). Both approaches aimed to identify the scope for potential savings company by company to determine the initial price and the value of X in an RPI-X type control in the coming period. Both approaches identified large differences in companies' operating costs. The average potential savings were 30 per cent (regression analysis)

and 24 per cent (consultants' study). The results and rankings of the two approaches were similar, both identifying the same companies as best (zero scope for improvement) and worst (both indicating scope there for 40 per cent or more improvement) (Offer, 1999).

As indicated, the comparative cost data are embedded in a system which has the central objective of providing incentives for efficiency improvement. Comparative competition can and does arise at all levels of regulation but the most powerful tool is that relative performance is taken into account at each Periodic Review. Companies which appear, from comparisons, to be operating inefficiently are penalised by not being allowed to increase prices by as much as the more efficient companies.

Section 2.2 highlighted that comparator competition helps to overcome asymmetries of information because the ratchet effect is weakened and there are disincentives for companies to report inflated cost estimates. Comparisons may reveal them to be inefficient and they will then suffer a tight price cap at the next regulatory review so that there is less benefit (in theory none) from hiding potential efficiency gains. Furthermore, the comparative data, much of which is in the public domain, provide information to others such as shareholders, analysts and customers who can also apply pressure to companies which appear to be inefficient to improve their performance.

The effects of benchmarking

There is strong evidence that the efficiency of the utilities has been improving since they were privatised. For example, in the water industry, a recent report found that operating expenditure fell by about 3.8 per cent a year in real terms between 1992/93 and 1997/98, while labour productivity improved by some 4.6 per cent per year from 1992 to 1997 (Europe Economics/Nick Crafts, 1998). However, the question we wish to address is what is the contribution of comparative competition and benchmarking since all of these efficiency benefits cannot be attributed to the comparative competition framework. The report concluded that the water industry had not made as much gain

in efficiency as other privatised utilities where stronger competitive pressures were present (Europe Economics/Nick Crafts, 1998). This would indicate that comparative competition might be an imperfect substitute for real competition even if we attribute all the efficiency gains to comparative competition.

This question as to the specific role of comparator competition arises every time there is a proposed merger of water companies since the number of comparators is reduced. The most detailed study arose in the proposed takeover of South West Water (SWW) in 1996 by Wessex Water (WW). Wessex's announcement of their intention to bid was followed by another bid for SWW by Severn Trent (ST). The Water Act 1991 requires any proposed merger of water enterprises to be the subject of a reference to the Monopolies and Mergers Commission (MMC) to see if the merger is in the public interest. Given that the proposed takeovers involved large water and sewerage companies and that three companies were involved, there followed a detailed investigation by the MMC, virtually all of which revolved around the value to the public interest of an additional comparator (since the merger would remove SWW as an independent observation for comparison).

The outcome of the investigation was very clear. The MMC vetoed the takeovers concluding that SWW is of substantial value to Ofwat for comparative purposes and that the loss of this comparator would weaken the comparative system across the range of uses to which comparisons are put. Most importantly the MMC stated that 'we take the view that in respect of this merger no remedy, even in the shape of very significant price reductions aimed at forcing the merged enterprise beyond the current efficiency frontier, would be sufficient to compensate for the loss of SWW as a comparator' (MMC, 1996). This is an extremely powerful statement as to the value of comparator competition to the public interest. If losing one comparator amongst the ten water and sewerage companies cannot be compensated even by the benefit of moving the new company beyond the efficiency frontier, then the public interest value of comparator competition where there is little real competition is large.

The MMC took the view that the loss could not be reliably quantified but Ofwat and SWW both presented estimated values. Ofwat estimated that the loss of SWW (poorly performing and the smallest water and sewerage company) amounted to £500 million in 1994/95 prices (approximately ten times SWW's current cost profit in 1994). Ofwat's estimate is the present value of the increase in costs that Ofwat thought would arise from the loss in ability to make comparisons between the companies if it lost SWW. This scale of loss arising from a reduction from ten to nine water and sewerage companies, gives a strong indication of the huge value that Ofwat attach to comparator competition.

Main lessons and problems

A striking feature of the experience of benchmarking costs in the utilities sector is the length of time it has taken for the system to settle down. Getting comparable data and a generally acceptable framework of analysis has taken a long time. For example, the water industry was privatised in 1989, but the comparative competition framework was still being adjusted and refined significantly in the 1999 price review. Similarly, Ofgem has recently embarked on a major 'Information and Incentives Programme' to improve the comparative incentive structure for electricity distribution companies.

The related issue of setting the appropriate benchmark has proved difficult. For example, regulators tend to analyse capital expenditure and operating expenditure separately when, in practice, they clearly interact with each other. This leads to several difficulties, most notably that by not using full cost approaches there is a real danger that the regulator identifies a false efficiency frontier. This happens because one is taking the lowest operating cost and mixing it with the lowest capital expenditure to identify what can be achieved, whereas in practice they are to some degree substitutable. The best performer on total cost may not be best at either operating or capital cost. This situation must change in the future.

There has been a real problem of knowing how far to push the

comparative analysis. A textbook model of benchmarking would set all prices from econometric modelling alone but in practice the regulators have used judgement. The degree of judgement can decline as data improves. This is confirmed in a comparison between the regime facing the electricity distribution companies and that facing water companies, which shows how much more econometrics can be conducted in the water sector because of the greater amount and homogeneity of data.

In addition to the time it has taken to identify the best benchmark, an interesting feature of the utility sector has been the time it is taking to identify the correct incentive structure. The traditional model of resetting prices at each periodic review to a level where the company is expected to earn a fair return in the future is now being dropped because of concerns over the effect on incentives. In the water sector, companies are allowed to keep benefits of capital saving for five years regardless of when they arise but it is not clear that the specific model used by Ofwat has solved the problems. The treatment of capital cost savings is also an issue of current debate in the electricity sector. This is proving a thorny issue because companies have followed very different strategies in terms of capital cost and operating cost savings. The present regime provides incentives to 'overspend' on capital relative to operating costs. Companies that have made significant capital cost savings are treated unfairly. The interesting lesson of this debate for this paper is the extent to which the incentive structure still remains problematic even after many years.

3.3 Central government

Experience to date

The main benchmarking initiative for central government is known as the Public Sector Benchmarking Project. This makes use of a model, the Business Excellence Model, originally designed for the private sector, to assess the performance of central government agencies across a range of business criteria. The project was launched in April 1996,

and has proceeded through three phases. Phase One was a pilot, involving 30 agencies, to test whether the model was applicable to public sector organisations. This was deemed a success. Phase Two sought to attract 35 agencies (participation was voluntary) but this was soon exceeded, and more than 100 organisations eventually took part. Phase Three was launched in April 1998, and will run for at least three years. The main distinguishing characteristic of Phase Three is that it is open to organisations across the public sector, rather than just agencies and governmental departments. The response has reportedly been impressive, with many organisations, including NHS trusts, the police, local government, and educational establishments, now participating in the project (Cabinet Office, 1999).

What is the problem that benchmarking is trying to address?
Benchmarking is part of a broad range of reforms of the public sector which, during the 1980s and 1990s, have attempted to replicate some of the competitive pressures existing in the private sector. The challenge has been to find ways of introducing a substitute for the profit-and-loss discipline which businesses face, and to change civil service culture to something closer to that prevailing in the private sector.

Since 1988 the Next Steps programme has made a clear division between the policy-making core and the rest of the sector engaged in the delivery of goods and services; the delivery side has been reformed by the creation of free-standing agencies which have been encouraged to become more business-like and less bureaucratic. Agencies have greater autonomy over such areas as the pay and recruitment of staff, with a weakening of central Treasury control. Performance related pay has been introduced into some parts of the public sector and agencies are set quantifiable targets annually.

There has, then, been a general drive for efficiency and value for money. Benchmarking can be seen as part of these changes. It introduces a management technique initially developed in the private sector, and represents a further attempt to improve performance within tight financial constraints. It also provides an opportunity for agencies

to learn from each other's good practice, as the Next Steps reforms have separated out the various parts of central government, and hence reduced the scope for ideas to flow between agencies. Benchmarking has the potential to counteract this tendency (Samuels, 1998).

Who is the instigator?

The Public Sector Benchmarking Project was instigated from the centre of government, the Cabinet Office. However, the decision of agencies to take part is voluntary. There is less compulsion in this sector than in education and local government, where benchmarking is imposed on organisations through league tables, or Audit Commission reports.

The underlying model in this case is one in which good practice is to be more widely disseminated within the public sector by the exchange of information between agencies. Although the scores of individual agencies are not made public, the database can be used to produce scores and charts for a range of criteria showing the agency how it performed relative to similar agencies. The agencies are also provided with examples of good practice which have been developed by various organisations to help them improve; they can search for a partner organisation which is also aiming to improve in a particular field. There are also workshops and conferences for managers of the public sector agencies.

The form of the benchmark

The Business Excellence Model measures business performance according to nine criteria: leadership, people management, policy and strategy, resources, processes, employee satisfaction, customer satisfaction, impact on society, business results. Weights are attached to each of the criteria, and a total score is given out of 1,000.

Given that the model was developed for private sector use, is it applicable to the public sector? The model gives a high weight to customer satisfaction while public sector organisations have to balance the needs of customers, staff, and taxpayers while delivering policies in line with priorities set by government ministers. The model assumes

that the organisation is operating in a competitive environment and points are scored for out-performing competitors. However, after pilot testing, it was decided that the model did have a good deal to offer public sector organisations.

The self-assesssment feature of the model raises some questions about the accuracy of the results. However, the conclusion reached on the accuracy of assessment was 'scores for the perception-based assessments in Phase One were considered accurate to within +/– 75 points, while those for the more rigorous assessments in Phase Two were considered to be within +/– 40 points' (Cabinet Office, 1999).

The main advantage of the Business Excellence Model is low cost. The total cost of Phase One (excluding staff time) was £75,000 +VAT or about £2,500 per agency involved. Most of this was the cost of the consultants employed on the project. It was hoped to reduce the use of consultants in the later phases as more civil service staff became trained in the use of the model. In Phase One the input required per agency was apparently a total of 35 staff days, spread among ten people and over six weeks, regardless of the organisation's size. This included materials, training and consultancy support. To put in place a benchmarking initiative across the whole of central government and beyond is clearly a major project, and alternative methods of doing so might have been much more costly.

Potentially, one use of the assessment results would be to examine how efficient the UK public sector is relative to the private sector. It seems that the public sector sample was below the average for the private sector on certain of the nine criteria suggesting that the UK public sector is inefficient. However, the database of 400 businesses used to give the private sector average is not representative. It over-represents high quality businesses (Cabinet Office, 1999). A comparison with the whole of the UK insurance industry suggests that the UK public sector was performing at a higher level on average than the insurance industry on every criterion except 'impact on society', a category which covered corporate responsibility, on topics such as green issues (Cabinet Office, 1999).

Perhaps a more useful exercise than the overall comparisons is to look at the relative strengths and weaknesses within the UK public sector. The scores achieved by the UK public sector agencies varied from below 150 at worst to almost 500 at best. The median score was around 350. This variation in scores implies that there is scope for agencies to improve by sharing good practice. The overall results indicate that, in comparison with the private sector users of the model, the strengths of the agencies included strategy development, good employment practice going well beyond the minimum statutory requirements, and sound financial management. The agencies were also well-focused on customer satisfaction, perhaps as a result of Citizen's Charter.

The role of incentives and targets

Over 90 per cent of agencies reported that one advantage of using the Business Excellence Model was that it had helped to improve their performance more quickly than might have been accomplished without the model. On average, agencies undertaking self-assessments found over 150 areas for improvement, and over 250 in some cases. Agencies claimed that the Business Excellence Model gave them a clearer picture of the inter-relationships between different parts of their organisation. The scope for comparisons helped them to realise what good practice really looked like, the Cabinet Office declared (Cabinet Office, 1999).

It has become common for agency chief executives to have performance related pay linked to the attainment of the targets. For instance, the Chief Executive of the Benefits Agency has the opportunity to obtain a bonus worth up to 15 per cent of her/his salary depending on whether the Agency has achieved targets relating to various aspects of its business including the accuracy and promptness of payments to claimants (National Audit Office, 1998). Many senior civil servants now have a performance element or merit award as a component of their salary, and performance related pay is becoming more common in the lower echelons of the civil service. For example,

the Inland Revenue has had a performance related pay scheme since the early 1990s, as have some other agencies, and there are plans to extend such schemes more widely.

We might therefore expect the results of benchmarking to show up as performance improvement against the targets set. If it could be shown that the benchmarking exercise had helped agencies to meet tough targets for improvement this would be compelling evidence of the relevance of the project. Agencies have met 75 per cent or more of their targets in each of the last five years but this may not mean very much unless the targets were challenging. In fact, many targets do not seem to be particularly challenging. Less than 30 per cent of targets in 1998/99 required agencies to improve on the performance level achieved in 1997/98. Performance improvements identified from the benchmarking exercise are not feeding through into tough targets for improving performance (Cabinet Office, 1999).

Moreover, the results of the benchmarking exercise are not in the public domain. This means that the pressures and incentives for weak organisations to improve their performance stemming from public scrutiny and public accountability which apply in sectors such as local government and education are not present in this sector.

The information produced by the Cabinet Office claims that the Public Sector Benchmarking Project has been a great success. These claims of success, however, are based mainly on the fact that a large number of organisations have chosen to use the Business Excellence Model and that these organisations assert that it has been helpful to them. The Public Sector Benchmarking Project may have the potential to improve the performance of central government agencies. But, at present, there is a lack of any hard evidence that it has actually done so.

3.4 Local authorities (excluding education)

Experience to date

Benchmarking developed gradually in local authorities during the 1990s. It has grown at the initiative of local authorities themselves,

and so has been patchy and incomplete in coverage. A number of benchmarking clubs have become established. The largest, the Inter-Authorities Group, has more than 80 members. There is a Local Government Benchmarking Reference Centre, a not-for-profit organisation which provides information to local authorities, and a range of other, smaller scale, initiatives.

In the last two years, however, benchmarking has grown rapidly among local authorities because of the government's Best Value programme. Performance measurement and performance comparisons are key aspects of Best Value. Although Best Value does not formally come into effect until April 2000, extensive pilot schemes have been running during 1998/99. A survey of over 300 local authorities conducted in 1999 found that 64 per cent reported benchmarking activity, an increase of 10 per cent on the previous year (Ball *et al*, 1999a).

Who is the instigator?

One major factor in the growth of local authority benchmarking was compulsory competitive tendering (CCT). This was one of the most significant and controversial initiatives of the 1980s. Under CCT local authorities could only continue providing defined services if they won the contract for them in open competition. The set of defined services was initially fairly narrow, concentrated on services such as refuse collection, but it was gradually extended to other services where it was much harder to write down contracts specifying appropriate terms and guarantees of quality. CCT posed a very real threat to the continuation of local authority run services. It provided a stimulus to benchmarking because managers in local government became aware that they lacked cost data with which to meet the challenge of market testing. Benchmarking clubs such as the Inter-Authorities Group emerged as a result (Davis, 1998).

A second factor in the development of benchmarking was the Audit Commission. The 1992 Local Government Act placed a duty on the Audit Commission to produce comparative indicators of local authority performance. It took some time to reach agreement on

what the indicators should cover and there was some resistance to the programme by local authorities. There were also concerns about the costs of collecting the data. The first set of figures related to 1993/94 and new data have been produced annually since then. The Audit Commission has adopted a neutral stance towards the data, letting the figures speak for themselves and not attempting to define good or bad performance (Cowper and Samuels, 1997). Nonetheless, the comparative data provided by the Audit Commission have put pressure on local authorities to instigate further work into areas of weak relative performance.

Benchmarking has also tended to be strong in services where there is an influential professional association, often developing from small-scale beginnings, by professional networks. An example is local authority housing, where many professionals are members of the Chartered Institute of Housing (Davis, 1998). Conversely, benchmarking has been at its weakest in services such as local economic development where CCT has been absent, where professional membership is weak, and where, in addition, there tends to be competitive rivalry amongst local authorities rather than collaboration.

As the reforms initiated by the present UK government become established, with the setting of national performance indicators, a requirement to undertake regular service reviews and engage in comparisons, and a duty to prove best value, it is likely that some form of benchmarking will become almost mandatory for local authorities.

What is the problem benchmarking is trying to address?

The government believes that local government is under-performing. The White Paper, *Modern Local Government: In Touch with the People* (DETR, 1998), claims, on the basis of Audit Commission performance indicators, that there are 'huge variations in service quality' and that 'such variations happen largely because of differences in council efficiency'. It further argues that only a few local authorities are of an excellent standard because local authorities generally lack adequate incentives (DETR, 1998, p 13). There is, then, an agenda for mod-

ernisation. Local government reforms are to be undertaken in order to press authorities to improve their performance.

Benchmarking is an important component of the reforms but its exact role remains unclear. Central to Best Value are the 'four Cs', which local authorities must use as tools, namely challenging, consulting, comparing and competing (Davis and Walker, 1998). Official statements on local government have sometimes envisaged the role of benchmarking as satisfying the comparison function. This implies that the emphasis would be on within-sector benchmarking, with councils learning from one another. At other times more stress has been placed on the competing function. In this model, external benchmarking would have to be used in order to prove that the local authority was at least as good as alternative service providers (DETR, 1998; Ball *et al*, 1999a).

The form of the benchmark

The Audit Commission indicators have been widely used in local government and the Audit Commission claims that they can be used to assess the economy, efficiency and effectiveness of local authorities. Some examples of these indicators are shown in Table 5. However, they have been criticised for being of limited relevance. Many are financial, concerning revenue or spending such as council house rent levels, and expenditure per capita on education. Such unadjusted data does not allow efficiency to be assessed. Other indicators reflect the Citizen's Charter emphasis on customer satisfaction, including response times for answering letters, the speed of answering the telephone, and procedures for handling complaints. These indicators cannot easily be used for comparative assessments because each authority sets its own targets, and definitions of what constitutes a complaint differ. Some indicators are based on such small numbers as to be of little or no statistical value. For example, maladministration judgements by the local authority ombudsman usually number only one or two for most councils (Boyne, 1997).

Local authorities themselves have been sceptical about the validity

Table 5 **Examples of indicators used by the Audit Commission**

Area of activity	Indicator
Education	% of 3 and 4 year olds with a local authority school place Expenditure per primary school pupil
Social Services	% of people aged over 75 helped to live at home % of children in local authority care who are in foster care Number of children on the child protection register per 1,000 children
Libraries	Number of books and other items issued by libraries per head of population
Housing	Average time taken to re-let local authority dwelling % of tenants owing more than 13 weeks rent
Planning applications	% of householder planning applications decided within 8 weeks
Council Tax benefit	% of new council tax benefit claims processed in 14 days
Crime and detection	Number of crimes recorded per 1,000 population % of all crimes cleared up by primary means % of burglaries cleared up by primary means
Police resources	Proportion of police officers' time spent in public Expenditure on policing per head of population
Fire Service	% of all fire calls at which attendance standards were met Cost of the Fire Service per head of population

Source: Cowper and Samuels (1997).

of the data produced by the Audit Commission, believing that problems of data comparability had not been adequately tackled (Stephens and Bowerman, 1997). Local authorities tended to see the indicators

as part of an agenda from the centre, attached to a culture of blame and shame, and with an emphasis on cost-cutting rather than quality improvement.

Under the Best Value reforms introduced by the Labour government a new set of performance indicators will provide information on both measures of the 'general health' of local authorities, and key indicators for each of their major services. Local authorities will be expected to set targets on the basis of these new indicators.

Best Value has increased the prevalence of benchmarking. A study for the Department of the Environment, Transport and the Regions (DETR) of 41 pilot authorities found that many had instigated new benchmarking exercises or else continued with initiatives that were already in place as part of their efforts to implement Best Value (Bovaird, 1999). However, they have either not been able, or have not even considered, the use of organisations in other sectors as appropriate benchmarking partners (Ball *et al*, 1999a).

Some authorities identified significant potential cost savings during the running of the pilot scheme (for example, Greenwich, as part of the London authorities benchmarking club, found that it was spending around £2-3 million more per year on domiciliary care than other inner London boroughs). The general impression, however, is that local authorities found benchmarking onerous. It was reported to be a 'slow and difficult process' (Bovaird, 1999).

A study of all 137 Welsh Best Value pilots found that only a minority were engaged in benchmarking prior to obtaining pilot status, and the others made little progress during 1998. There were difficulties in identifying benchmarking partners and problems in persuading other authorities to join benchmarking clubs, either because of concerns about obtaining confidential data or because non-pilot authorities were simply not interested. Other difficulties included the development of common methodologies for costings or the collation of performance data, and problems of data interpretation (Boyne *et al*, 1999).

The role of incentives

In the new system, the lure of 'beacon' council status will provide an incentive for some high-performing local authorities. They are an important part of the present government's reform programme. Beacon councils are the very best performing local authorities. They will be selected on the basis of excellence in service delivery against national and local performance indicators and targets. Local authorities can apply for beacon status for specific service areas or for the local authority as a whole.

Those councils with beacon status for certain services will have more freedom to make capital investment in that service, and legislative controls on the local authority may be eased for that particular service.

Councils with overall beacon status will gain additional powers and freedoms. These could include being exempted from central government powers to cap council tax increases. They may also be freed from some statutory constraints on service delivery and they may be allowed to levy, within certain limits, additional business rates (DETR, 1998).

At the other end of the scale, it is intended that under-achievement be rooted out by a new system of audit and inspection. Among the duties of the Best Value Inspectorate are to ensure that performance reviews have been carried out and that targets are challenging. Central government has powers of intervention including: requiring an authority to draw up an action plan for improvement, requiring an authority to accept external management help, and putting services out to competition (DETR, 1998).

The incentives and penalties for the very best and the very worst local authorities are, therefore, clear enough. What is less clear are the incentives facing the bulk of councils in between these two extremes. Under the Best Value framework, authorities will be freed from the pressures of CCT but will have an obligation to carry out comparisons of their performance. It remains to be seen whether performance comparisons are enthusiastically taken up by local authorities generat-

ing lots of new benchmarking activity, or whether local authorities do the minimum that they can get away with.

The effects of benchmarking

Whether benchmarking in local government has been of any value remains 'indeterminate' (Davis 1998, p 268). There has been little or no evaluation of its success or failure. The kind of benchmarking undertaken for Best Value has so far concentrated on within-sector comparisons (Ball *et al*, 1999a,b). Local authorities may be fearful of benchmarking against external competitors because they do not want to lose services as under the old CCT regime. Concerns have also been raised that central government is trying to impose its own agenda accompanied by a 'name and shame' approach, and a continuing emphasis on fiscal restraint. Such a framework might work to prevent more creative use of benchmarking as a tool for locating and implementing best practice.

Underpinning this ongoing debate is a lack of clarity about the nature and extent of the problem that benchmarking is trying to address, and therefore the appropriate system of incentives that should be in place. Benchmarking as an exchange of best practice has been the predominant pattern amongst councils to date. On the other hand, some of the government's pronouncements seem to be predicated on a version of the alternative model where there are poor incentives. However, the government has not been consistent. More discussion and analysis is needed in order to determine which of these alternative approaches to benchmarking is really the best one for local government.

3.5 Education

Experience to date

Benchmarking in education has taken the form of league tables of exam results. The Parents' Charter of 1991 required schools to provide such information. Comparative data on the results for GCSE exams

and 'A' levels (the major qualifications available to school students, at ages 16 and 18 respectively) were published for the first time in 1992. More recently, primary schools have been required to publish the results of standard assessment tests on their pupils. This information is intended to aid parents in their choice of schools, and to goad schools into improving their standards.

Criticism of league tables has centred on the use of 'raw' exam results as the measure of performance (e.g. Goldstein and Thomas, 1996; Higgs *et al*, 1997). Such use made no allowance for the ability of pupils on entry to the school. A school which was able to select very able pupils might achieve excellent results in the league tables, but this did not necessarily mean that it was providing better quality education. Conversely, a school which took in mainly pupils with disadvantaged backgrounds could greatly improve the attainment levels of its pupils yet might still be placed in a lowly position in the league tables.

In 1995, the then Conservative government launched a study into the design and piloting of a national system of value-added measures (Saunders, 1999, p 235). This move towards the use of value-added information has been continued by the present Labour government. The White Paper, *Excellence in Schools*, published shortly after the 1997 general election, confirmed the government's commitment to value-added performance data, and in 1998 a large pilot of value-added for GCSE results was conducted. National value-added information for GCSEs will be published in 2000; a working party in the Department for Education and Employment (DfEE) has been established to examine value-added at 'A' level, and value added measures at primary school level are expected to be introduced by 2003. Value-added information will be published alongside the raw exam scores.

Who is the instigator?

The introduction and use of league tables has been driven by central government. The league tables were part of a broader package of reforms, which were intended to introduce a quasi-market into education.

Under the Education Reform Act of 1988, provision was made for local management of schools (LMS), so that funding was devolved from local education authorities to schools, to be managed by the head teacher and the school governors. If they wished, schools were allowed to opt-out entirely from local authority control, and receive funding direct from central government (Glennerster and Hills, 1998). Among other measures introduced by the 1988 Act was a provision for allowing school catchment areas to overlap, so that parents had some degree of choice of which school their child went to. Prior to this pupils had been allocated to a particular school on the basis of home address (Higgs *et al*, 1997). Important changes to school funding were also made, with the introduction of formula funding so that the money a school received was more closely linked to the number of pupils it attracted (Bradley *et al*, 1999).

Not all the reforms have pointed clearly in the direction of decentralisation and marketisation. The National Curriculum, also introduced in the 1988 Act, reduced the autonomy of teachers, and the system of school inspection has become much tougher with the creation of the Office for Standards in Education (Ofsted). Similar reforms to state schools have been made in several other countries, including New Zealand and the United States. Although the reforms have the avowed intention of deregulating, introducing quasi-markets, and devolving managerial responsibilities to schools, in practice they have often tended to be accompanied by increased central government intervention (Gordon and Whitty, 1997; Power *et al*, 1997).

What is the problem benchmarking is trying to address?

The problem is to raise standards in education. The government believes that, whilst excellence occurs at the top of the education system, many children under-achieve at school. Two-thirds of 16 year olds do not achieve GCSE grade C in both maths and English. International comparisons suggest that the UK is well down the table in terms of attainments in maths and science, and relatively low proportions carry on to university education (DfEE, 1997).

The underlying model is one in which it is assumed that incentives in state schools are weak, and a root cause of poor performance. The absence of incentives, it is argued, has enabled poor teaching to go unpunished and encouraged complacency. League tables, and other quasi-market reforms, apply pressures to schools which lead on to better performance and higher standards (DfEE, 1997, p 25). Parents would be able to tell from the exam league tables which schools were good performers and which ones bad, and choose the best school for their child accordingly. The ability of parents to choose meant that schools would have to compete for pupils. Successful schools would attract more resources, via formula funding, and weaker schools less, providing incentives for schools to perform well.

The form of the benchmark

The comparative measure chosen for schools is based on exam results. The main disadvantage of the measure is that it takes no account of contextual factors which influence exam performance.

There are, however, also problems associated with a value-added system. A focus group study by DfEE found that very few parents had heard of value added, although they welcomed the use of the concept when it had been explained. However, when shown an example from the pilot scheme on value added, they found the data far too complex and difficult to understand (DfEE, 1999).

The data requirements for a value-added measurement system are high. The system being developed in education require data on individual pupils, tracking them as they move from school to school (DfEE, 1998). This helps to explain why the transition to value-added measures is taking several years to complete.

There are different ways of measuring value-added in education. The government has chosen to adjust raw exam scores to allow for pupils' previous attainment on entry to the school. This has been justified by reference to a study from the School Curriculum and Assessment Authority which apparently showed that prior attainment was the best predictor of exam scores regardless of other factors. Using

other variables would add to the complexity of the system (DfEE, 1998).

Other issues also arise:

- exam scores, whether computed on a raw or value-added basis, are only a single measure of outcome. But parents, in selecting a school for their child, will be swayed by a range of other factors besides its aptitude for getting pupils successfully through exams;
- the league tables do not take account of the relative performance of different types of pupil within a school. Some schools may be very good with advantaged pupils, and not so good with disadvantaged pupils, or vice versa;
- results may not be reliable over time and across subjects. Thomas *et al* (1997) looked at data on students' GCSE exam results from 94 inner London secondary schools across three cohorts, 1990 to 1992. A value-added aproach was utilised. The main conclusion was that very few schools performed both consistently (across subjects) and with stability (over time);
- performance measures reflect the past not the present. In the case of schools there is a long lag, as exam results of 16 year olds will depend to some extent on teaching which they received up to five years previously. Schools may have altered considerably over this time.

League tables therefore convey imperfect information to parents considering where to send their child.

The effects of benchmarking in education

It must be borne in mind that the league tables were part of a package of reforms, making it difficult to disentangle the effects that they have had from the influence of other changes. That said, we focus on two broad aspects of the likely effects of the league tables. Have they provided useful information for parents? And, what effect have they had on schools?

The league tables are certainly being used by parents to make choices about schools, and there seems to be a tendency for their use to increase over time. In three areas studied by Woods *et al* (1998),

responses to questionnaires showed that roughly a fifth of parents stated that they had used the league tables as a source of information, with an increase over the period 1993 to 1995. Parents also obtained information about the schools in their area from a range of other sources, including visits to schools, personal information on the schools, other parents, school brochures, and siblings at secondary school. A range of other factors were frequently cited as having a major influence on school choice. These included nearness/convenience for travel, the child's own preference, the fact that a child's friends would probably be going to a particular school, the fact that an elder sibling was already at the school, the facilities provided by the school, and the caring approach of the school (Woods *et al*, 1998).

However, a significant proportion of schools are over-subscribed, so that children cannot be allocated to the school of their parents first preference. The most popular schools have not expanded their capacity sufficiently rapidly to be able to deal with the demand for places. A 1996 Audit Commission report found that as many as one in five parents had not been able to obtain a place for their child at their first choice of secondary school (cited in Taylor and Bradley, 1998).

Have the reforms been successful in raising standards in education? The proportion of pupils attaining five or more good GCSE passes has certainly increased during the 1990s (DfEE, 1997) but this could be because the exams have become easier, rather than because pupils are getting a better education.

There is some evidence that competition among schools has increased. Hardman and Levacic (1997) studied the relationship between the success of schools in the market, financial success, and their exam performance. The analysis was based on a sample of some 300 English schools, from six local education authorities, chosen to obtain a spread of rural and urban areas, and to reflect diversity of socio-economic backgrounds and school systems. They developed a market success indicator (MSI), which was defined as the ratio of the actual intake of pupils to the local cohort size. Those schools in which the MSI was increasing between 1989/90 and 1993/94 or capacity

had been reached were classified as 'improving/full'; schools in which the MSI was consistently falling over time, or fell off and then remained steady, were categorised as 'declining/plateau'; other schools, in which neither of these classifications applied, were referred to as 'middling'. The main result of the study was that the 'improving/full' schools tended to have a higher proportion of their pupils achieving five or more GCSE passes at A-C grades than 'middling' schools, which in turn tended to do better than the 'declining/plateau' schools. These findings suggest that parental choice among schools on the basis of exam results was occurring. However, the authors also discovered that the government's claims that extra funding would flow to those schools which were successful in attracting students were not being fully borne out in practice.

Bradley *et al* (1999) analysed a large dataset on over 2,500 secondary schools which served their local areas and did not select pupils on the basis of exam or interview over the years 1993 to 1998. They measured the extent of competition faced by a particular school by counting the number of other secondary schools within a specified distance. Data envelopment analysis (DEA) was used to assess the 'efficiency' of schools according to the proportion of pupils obtaining five or more GCSE passes at grade C or better, and the truancy rates experienced by the schools.[3] Controlling for the type of school and for the socio-economic background of the area in which the school was situated, they found that schools which faced most competition tended to be more efficient. The extent of competition was also found to exert a positive effect on the change in 'efficiency' over time i.e. the greater the extent of competition between schools, the more 'efficient' schools tended to become.

Woods *et al* (1998), reporting the results of a study of three contrasting areas found that competitive pressures had intensified in two urban areas, but had had little effect on their most rural area. As competition increased, so co-operation between schools declined. Co-

3 DEA is discussed, in the context of utilities, in Section 3.2 above.

operation between schools was strongest in the semi-rural area where competition was weakest. Here, the heads of schools observed an informal, working agreement about which local primary schools they carried out promotional activity in, and did not poach from each others feeder schools. Such co-operative relations no longer existed in the more competitive areas.

What other effects have the competitive pressures had on the behaviour of schools? In response to the educational reforms, schools have focused more attention on promotional activities (Gewirtz *et al*, 1995; Woods *et al*, 1998). Substantive change to the teaching in the school is less in evidence. Nonetheless, some changes have occurred. One of the schools studied by Woods *et al* (1998) had placed more emphasis on its strengths as a technology school. Some schools were also keen to attract pupils by offering plenty of options.

It is well known that performance indicators will often have unintended side effects. This arises from the multi-tasking nature of teaching, as discussed in Section 2.3 above. These side effects can be classified into two broad categories: firstly, focusing on certain aspects of performance at the expense of others, and, secondly, manipulating the signal of performance either by altering the data in some way, or else by engaging in strategic behaviour (Smith, 1995). There is evidence of both these effects in the education system. In particular:

- the pressures to perform well in the league tables have encouraged schools to engage in cream-skimming activities (Woods *et al*, 1998; Adnett and Davies, 1999). Cream-skimming means discrimination by providers of a service against more expensive users. Schools were keen to attract a higher proportion of applicants who are likely to do well at GCSE, and adjusted their image to appeal to the parents of such children;
- schools have become less willing to take pupils with special educational needs because they were unable to gain extra or sufficient resources beyond the funding formula for children in this category (Evans and Vincent, 1997). Schools catering for special educational needs receive some additional funding, but this appears to be insuffi-

cient to allow for the additional costs incurred;

- there is evidence that disadvantaged pupils who are unlikely to score good exam passes and less academically able students have been discouraged from entering for exams; the number of permanent exclusions rose from about 3,000 in 1990 to over 13,000 in 1997/98 (*Times*, September 17th 1999) and this may be related to the increasing pressure on schools to perform well in league tables;
- concerns have been expressed that teachers will give disproportionate attention to pupils close to the grade C/D borderline at GCSE and may, therefore, neglect other pupils. Hard evidence on this is difficult to find, although the National Association of Head Teachers has voiced concern that it has been happening (*Guardian*, August 26th 1999).

Overall, then, it is likely that the league tables, combined with other reforms, have had some effect in raising competitive pressures on schools. The extent of competition has varied considerably, because state schools operate in local, rather than national, market places. There are, however, some serious concerns about the equity effects of the reforms. The multi-tasking issues, discussed in Section 2.3, have been much in evidence. Woods *et al* (1998) concluded that, in the more competitive areas they studied, there had been a gradual tendency to lay more emphasis on academic success as the main criterion of good schooling. This was a subtle process, and did not mean that other aspects of child development had been neglected by schools. Nonetheless, there had been increasing emphasis, in the schools studied, with concentrating time and effort on policies and practices which were felt likely to influence measured academic achievement. The personal, pastoral and social aspects of schooling had tended to receive a reduction in emphasis, although they were still important.

3.6 Summary

From this review of experience across different sectors, it is clear that the nature of benchmarking differs from sector to sector in terms of

the problem that is being tackled, the clarity with which the problem is addressed, and the extent to which incentives are in place. We now draw together the main points which emerge from our analysis.

There is evidence from several sectors that the underlying problem that benchmarking is designed to tackle is often not thought out clearly or made explicit. For example, in local government, there appears to be genuine confusion and mixed messages as to whether benchmarking is about raising average performance through the use of incentive mechanisms, or about the diffusion of good practice. In the case of education, little or no attention has been paid to schools that face very limited competition. Nor have the side effects of incentivising the education sector received much attention.

The extent to which incentive mechanisms are in place varies. In the private sector, the profit motive provides an incentive for firms to eliminate areas of weakness and to seek out best practice. In the utilities sector, comparative cost data are embedded in a system that aims to provide incentives for efficiency improvement by mimicking the effects of a competitive market. The incentives in the public sector cases are less transparent. For central government agencies, the motivation for agencies to raise performance is not clear. In local government, although there is now an embryonic system of incentives to raise performance through the beacon councils initiative, this is likely to apply only to the very best local authorities. In education, the incentives to improve performance are clearer: schools earn extra funding by attracting pupils. Yet high-performing schools do not expand sufficiently rapidly to meet demand, thereby obstructing the parental choice that the system is designed to promote.

If benchmarking is instigated from the centre one danger is that it will antagonise organisations, which will feel that they do not own the benchmarking initiative. On the other hand, if benchmarking does not emanate from the centre, then the coverage of benchmarking is almost certain to be patchy and incomplete, especially in parts of the public sector where there do not exist strong incentives to improve organisational performance.

There is a widespread belief that benchmarking is more prevalent and of a more sophisticated form in the private sector than in the public sector. In fact, the general picture is that, while benchmarking may have been more prevalent in the private sector than the public sector in the 1980s, this is probably not true today. Benchmarking, in one form or another, is now common in many parts of the public sector. It is seen in some cases as a substitute for competition and in other cases as a means to stimulate choice by uses of public services. Although many large firms make use of benchmarking, there are barriers in terms of knowledge of the technique and the level of resources required that have prevented benchmarking being used by small and medium-sized enterprises.

Nor does it seem that private sector benchmarking is of a more sophisticated form than in the public sector. Both sectors report problems in moving beyond relatively straightforward comparisons of performance towards utilising benchmarking fully as an effective tool for improving performance. It should not, then, be assumed that the public sector lags far behind experience in the private sector.

There is a worrying lack of evaluation of benchmarking in most sectors except education. It is difficult to find clear evidence that benchmarking activity improves the performance of organisations, still less that it raises performance sufficiently to offset the costs of implementing benchmarking. This finding appears across several of the sectoral case studies, including local government, central government, and the private sector. It is partly a consequence of the nature of the literature on benchmarking, much of which is more concerned with promoting than appraising. In some parts of the public sector, such as central government agencies, it reflects the relative recentness of its introduction. But it also relates back to one of our main themes in this paper, that the problem that benchmarking is addressing has often not been clearly thought out.

4 BENCHMARKING AND THE NHS

4.1 A brief review of experience to date

The development of benchmarking

There has been a long tradition of utilising comparative data in the NHS. A widely used set of benchmarks was the English NHS Performance Indicators. These were established in 1983 and provided comparative information on over 100 indicators of performance, reported for each health service district. The objective was to provide information to hospital managers to improve their performance. Subsequently, the number of indicators was increased and the indicators were distributed in computerised form, with the government developing an expert system to enable the user to analyse the data.

With the arrival of the NHS internal market in 1991, the data were renamed the Health Service Indicators and the focus shifted to the performance of hospitals and other providers, with the intention that comparative information would be available to purchasers of health care to help them with their decision making (Nutley and Smith, 1998). With the advent of the NHS internal market came more indicators of comparative performance. The Purchaser Efficiency Index was a Department of Health set target for hospitals to increase output by three per cent per annum at a constant level of costs. The Purchaser Efficiency Index was criticised for providing incentives for hospitals to increase hospital-based activity relative to other services, and for diverting attention away from obtaining improvements in the quality of health care (Dawson and Street, 1998, p 16). The Patient's Charter provided information on the comparative performance of hospitals from 1992 onwards. The comparative information included within the Patient's Charter included measures of: percentage of outpatients seen within 30 minutes of appointment time; percentage of accident and emergency patients assessed within five minutes of arrival; percentage of outpatients seen within 13 weeks of referral; percentage of inpatients admitted within three months of being put on a waiting list; percentage of patients admitted within 12 months of being put on a waiting list (Nutley and Smith, 1998).

Almost all the measures used in assessing performance to date have been either costs or measures of throughput. There has been less focus on outcome measures, though data on clinical outcomes has been available publicly in Scotland since 1994. However, in 1999 the Department of Health developed the High Level Performance Indicators (HLPIs).

The High Level Performance Indicators

These cover six areas: improving people's health, fair access to services, delivering effective health care, efficiency, the experience of patients and their carers, and health outcomes. Several performance indicators are included under each of these headings, making a total of 41 indicators. There are also six clinical indicators of patient outcome which cover a limited range of specific aspects of clinical care including deaths in hospital following a heart attack, or after surgery for a fractured hip.

The indicators of health improvement include deaths from all causes, from several types of cancer, from circulatory diseases, and from suicides and accidents. Indicators of fair access include surgery rates for common operations (knees, hips and cataract replacements), size of waiting lists and adults and children registered with an NHS dentist. Measures of effective delivery of health care include percentage of the target population vaccinated and screened for breast and cervical cancer, rates of inappropriate surgery, surgery rates for common operations, the management of chronic care, and measures of appropriate prescribing. Measures of efficiency include the day case rate, casemix adjusted length of stay, unit costs for patients in mental health services, and the percentage of generic prescribing. Measures of patient and carer experience of the NHS relate to waiting times, cancelled operations, and delayed discharge for the elderly. Measures of outcomes of NHS care include conceptions below age 16, treatment of teeth in five year olds, emergency admission rates, potentially avoidable mortality measures, and measures of premature deaths in hospitals.

This is clearly a wide ranging set of indicators and measures the activity of health professionals in all parts of the NHS. The focus on outcomes and patient experience moves the focus away from the more traditional focus on inputs and throughputs. The wider range reflects the broad aims of the NHS. Some measures are clearly more under the direct control of clinicians and other professionals in the NHS than others: deaths from suicides and teenage conceptions are clearly less easy to influence than generic prescribing or surgery rates for common operations. Further, it is clear than some of the measures will be more affected than others by the socio-economic status of the population. These issues are recognised by the Department of Health in their discussion of the indicators.

The indicators are new and therefore there is little experience of their use to date. There is however more experience in the use of reference costs and we now focus on this.

The use of reference costs

Most effort to date has focused on measures of cost, and here we concentrate on the latest measures of these. We do not discuss here in detail the technical issues in their construction (for further information on this see Street, 1999) but outline the type of problems that have arisen in their construction and in their possible use for improving performance.

In November 1998 summary data of NHS Trust costs were published in the form of the Reference Cost Index (RCI). The RCI is a weighted average of all Healthcare Resource Group (HRG)[4] costs in each Trust relative to the national average, adjusted by the Market Forces Factor (MFF) to take account of differences in local factor costs such as higher labour costs in the London area. There have been a number of criticisms of these measures. The first is that they do not

4 Healthcare Resource Groups (HRGs) were designed as a management tool. Each HRG consists of groups of patients who are expected to consume similar amounts of health care resources.

allow for many of the factors which have an influence on hospital costs but which are beyond the control of management. This has led to the development of the CCI measures (see below). Second, the quality of the data is sometimes in doubt. Difficulties include whether the use of HRGs adequately reflects casemix complexity and whether financial cost measures accurately reflect resource use. In their discussion of reference costs, Dawson and Street (1998) show that the quality of cost data in the NHS is well below that in many other countries. In particular, there is a lack of patient-related costing information. Dawson and Street have also argued that there may be small number/statistical inference problems associated with the use of HRGs. It has been estimated that some speciality HRGs at individual Trust level contain between one and nine cases per year. Cost estimates based on such small samples are not statistically robust.

Many of these criticisms are very similar to those levelled against benchmarks used in the utilities and in the rest of government. Picking the appropriate comparator is difficult, and poor quality data and/or misrepresentation of the data are common in the development of performance indicators and benchmarks (Smith, 1995).

In response to acknowledged problems with the existing measures of performance, a new set of casemix cost indices was developed by the Department of Health and the Audit Commission early in 1999. These measures attempted to take into account those factors that cannot be controlled by management, so that a measure of their true efficiency can be constructed. As with the cost analysis developed by Ofwat in the water industry, the indices use regression analysis to factor out the exogenous variables, leaving a residual that reflects the relative efficiency of Trusts. Three separate but related indices have been constructed which are known as CCI, the Casemix Cost Index; 2CCI, the Casemix Costliness Index; and 3CCI, the Casemix Costliness and Configuration Index (Street, 1999).

The CCI cost index takes into account case mix as measured by HRGs, day case, inpatient, first outpatient attendance and accident and emergency activity. It is a ratio of actual to expected costs, where

expected costs are taken as the national average cost. The 2CCI cost index allows for a range of additional factors, including further adjustments for casemix, age and gender, transfers in and out of hospital, inter-specialty transfers, local labour and capital prices, and teaching and research costs. In practice, variations in efficiency on the 2CCI may arise because of factors beyond the immediate control of management. The 3CCI attempts to allow for some of these factors, including the costs of multi-site working, hospital size, and capacity utilisation, by including data on number of beds, number of sites, the scale of inpatient and non-inpatient activity and the scope of activity. The 3CCI is, then, a short-run index. It takes account of hospital capacity which may be fixed in the short run, but amenable to change in the longer term, while the 2CCI is a long-run cost index.

Evidence on responses to these measures

Most of these indices are at the development stage, so the relevant question is how will Trusts respond to league tables of cost data? As is clear from our discussion in Chapter 2, faced with the introduction of a comparative benchmark, a Trust which has high costs relative to others could respond in several ways, some beneficial, others not. Consider the simplest situation where there is only one activity that is being measured. First, the Trust could put in more effort and improve its outcome. Second, it could set about improving the quality of the data. This would mean investing in more administrative support to improve its coding and making its cost allocation more accurate. This would be a desirable outcome, leading to improved data quality in the NHS. However, both responses one and two could involve diverting resources from achieving other outcomes that were important but not being benchmarked. Third, it could use resources to re-classify cases to enhance its position in the league table without real efficiency improvements. Dawson and Street (1998) note that there is evidence of this occurring in the US following the introduction of DRG funding. Fourth, it could react to a poor league table position by reducing length of stay and increasing throughput. The risks of this strategy

include the shifting of costs onto other care agencies, such as social services, and reductions in the quality of care. Fifth, the Trust could take no action at all.

In a survey which looked for possible unintended consequences of the introduction of performance indicators in the NHS, Goddard *et al* (1998) identified two types of consequence. First, performance indicators can encourage managers to concentrate on certain aspects of performance and neglect other important aspects, rather than focusing on performance in the round. Here Goddard *et al* distinguish between:

- **Tunnel Vision:** Concentrating only on the areas that are covered by the performance indicator scheme.
- **Sub-optimisation:** The pursuit of narrow objectives by managers at the expense of the objectives of the organisation as a whole.
- **Myopia:** Focusing on short-term issues at the expense of the long run.

Second, performance indicators may lead to manipulation of the signal of performance. Goddard *et al* (1998) distinguish between:

- **Misrepresentation:** Fiddling the data through creative accounting or fraud.
- **Gaming:** Altering behaviour in order to obtain strategic advantage[5].

Goddard *et al* (1998) report the results of interviews with chief executives and medical staff in eight NHS Trusts on whether the performance measurement system was having these kind of dysfunctional effects. *Tunnel vision* was found to be common, notably through the importance given to meeting waiting time targets which was felt to divert resources away from other areas. Specific examples cited included treating all cataract patients on the list as a means of reducing num-

5 In terms of the discussion in Section 2.3 above, the first three responses are examples of responses to performance indicators in the presence of multi-tasking and the second two are possible responses to the introduction of benchmarks with associated rewards for good performance and penalties for bad performance.

bers at a relatively modest cost. Also, the employment of 'hello nurses' in hospital accident and emergency departments who contact patients on arrival in order to ensure that the five minute waiting time target was met. Concern was also expressed that the lack of measures of outcome meant that managers focus on the many measures of process.

Sub-optimisation can arise in various forms including the lack of congruence between the financial objectives of the Trust and the objectives of the clinicians working within it; also other agencies involved in care, such as social services, may have differing objectives or priorities to the Trust. *Myopia* did not seem to be a serious problem to the people interviewed, although it sometimes arose e.g. Trust chief executives moving on before severe problems hit the Trust.

There was a mixed picture on the question of *misrepresentation:* some respondents giving examples of double counting of finished consultant episodes (FCEs) when a patient is referred to another consultant within the same hospital, while others maintained that it did not occur in their organisation. For *gaming* the most common example was the Purchaser Efficiency Index: respondents stated that they were not keen to obtain large gains in a particular year for fear that they would be expected to achieve similar gains in future.[6]

More generally, there is an issue of whether the current benchmarks have any impact. The York Health Economics Consortium (YHEC, 1999) survey asked Trust managers in their sample where unit costs were above the mean whether they would be likely to take action, specifically making contact with good practice sites on the basis of the information contained in the new cost indices (CCI, 2CCI, 3CCI). It was found that very few sites intended to take action. A range of reasons was reported. Some managers felt that the data were simply incorrect: they did not take account of local circumstances, or

6 The US DRG system of payments has been argued to have severe problems of misrepresentation and gaming; an example being so-called 'DRG creep' where patients are given treatments which give the organisation more income even if those treatments are not strictly medically necessary.

differences in cost apportionment. Other managers reported that the data did not tell them anything new and that action had already been taken or was being considered. Some managers accepted that they had high costs but pointed to factors such as a complex casemix, even though the data had been adjusted to reflect this. The problem of getting clinicians on board to make necessary improvements to processes was seen as a major headache in some Trusts which accepted that they had high costs. Many Trusts complained about the existence of both Reference Costs and the new indices, maintaining that there should be just one set of cost indicators.

Many directors of finance thought that, unless funding was based on national standard costs at specialty and procedure level, then the cost indices would be regarded as 'interesting but lacking in bite'. There were no incentives to take the numbers seriously. Health Authorities focus on costs for the Trust as a whole. High costs in one specialty are acceptable so long as they are offset by low unit costs elsewhere. A Trust would only focus in detail on costs at specialty level if it was under pressure from a Health Authority about costs in the aggregate. It appears that the main objective for Trusts is to stay in the 'comfort zone'. So long as they do not sink to the bottom or near bottom of the league table, questions will not be asked of it and there will be no real pressure to improve.

The relevance of these responses

These responses are not surprising. As outlined in Section 2.3 above, where agents engage in multiple tasks (as all Trusts and Health Authorities in the NHS do), the introduction of comparative measures of performance which are linked to rewards and sanctions are likely to have unintended side effects. The central problem is that of asymmetry of information: the person or organisation designing the performance measures (the principal) has less information that the person or organisations (the agent or agents) it wishes to monitor. This means that the agent can take actions in response to measures of comparative performance that the principal might not wish. The problem is likely

to be compounded where there are multiple principals and where those who are being monitored have career concerns that do not align exactly with those of the organisation they work for. For a fuller discussion of these issues in general, see Section 2.3.

All these factors are likely to be a problem in the NHS. Almost all organisations within the NHS, from the Department of Health down to Trusts and Health Authorities have multiple tasks. Many have several principals whom they serve. For example, the Department of Health is charged with delivering high quality health care whilst staying within a tight annual budget. It responds to politicians, voters, special interest groups, labour unions and its own civil servants. Trusts have outputs that include patient care, teaching and research. They answer to the centre (the Department of Health and the NHS Executive), to their Health Authority and Primary Care customers, to patient groups and to the local population. Primary Care Groups/Primary Care Trusts (PCGs/PCTs) answer to the centre, but face pressures from patient and other consumer representatives, local politicians, and professional bodies. The same set-up is replicated within each of these organisations. For example, within a Trust clinicians engage in multiple tasks – teaching, research, patient care are three broad groups of tasks. Clinicians are line managed by the Trust chief executive, but are motivated by career concerns which are not necessarily the same as the goals of the Trust. Clinicians may value undertaking research to be published in academically prestigious journals or participation in the work of their professional societies more than the manager of a Trust who is concerned with delivering patient care within a fixed budget. Other groups within Trusts, and clinicians and medical staff in PCGs/PCTs also have multiple tasks.

We now consider the implications of the nature of the NHS and the organisations within it for the design and use of benchmarks in the NHS. We focus on the three issues identified in Chapter 2: the need to address exactly the problem that benchmarking is intended to solve, the form the benchmark should take, and the strength of the link between the benchmarks and financial rewards.

4.2 What is the problem benchmarking is trying to address?

Benchmarking in the health service has been proposed to meet a large range of objectives. At their simplest, benchmarks are promoted as helping managers and clinicians know their performance, helping identify poorly performing units. They are also advocated as a method of helping purchasing or investment decisions. Best practice models have also been used in health. And benchmarking is used to convey the centre's priorities and 'vision for the NHS'.

Commonly cited uses for the benchmarks proposed for use in the NHS at present include (Richmond House Workshop, 1998; YHEC, 1999):

- Trusts using the database of benchmarking information to investigate differences in relative costs at both the Trust and specialty level;
- Trusts using benchmarks to compare their practice and exchange ideas/interchange of ideas with suitable Trusts elsewhere in the country;
- Health Authorities and PCGs/PCTs using benchmarking information to compare the performance of providers from whom they are commissioning services;
- Trusts, Health Authorities, and PCGs/PCTs using benchmarking information and associated indicators as part of their decision criteria when making investment decisions;
- The centre using benchmarks to identify poor and good performers.

There appears, however, to be less clarity over the exact nature of the core problem that benchmarks are supposed to address. As noted in our discussion of theory in Chapter 2 there are two possible polar cases. The first is that health care providers are similar in terms of the effort they supply, but lack incentives to provide effort because the public sector is poor at giving incentives for efficiency. Under this view of the world the overall level of efficiency is too low and benchmarking is needed to bring up everyone's efficiency. In doing so the average

level will also be raised. The second is that health care producers differ considerably. There is considerable variation in behaviour: there are some outstanding producers, and some who are poor. Producers are motivated by career concerns, but the labour market is not sufficiently active for good practice to be disseminated through hiring, and poor practice reduced through firing. Under this view of the world, it is not that there needs to be great incentivisation of the sector, but that good practice needs to be disseminated so that it can be copied and the practice of poor performers raised.

In reality, the world is likely to be a mixture of these two polar cases. Poor incentives may be a problem in some areas, poor dissemination of good practice a problem in others. Applications of benchmarking in health care embody both views. The first view underlies the use of DRGs in the US health care system. For each DRG an average cost of treatment is calculated. The payment made to any health care provider who supplies this treatment is this average cost. Suppliers whose costs are lower than the average will make a profit. Conversely suppliers whose costs are above the average will make a loss. The assumption that underlies this mechanism is that within DRG suppliers differ only in efficiency, and that efficiency is under their control. The incentive scheme is high powered: the gainers are those who are efficient, the losers are those who are inefficient.

The second view led to the widespread use of 'good practice' models in local government (see Chapter 3 for further information on local government). In support of the second view in the NHS, it is clear that NHS managers are motivated by career concerns. In the case of clinicians, the importance of the esteem of their peers is high. Consultants are judged by their performance relative to others in their specialty, and their pay is set not in relation to their colleagues within their hospital, but in relation to others nationwide in their field. Clinicians in PCGs/PCTs are similarly motivated by career concerns.

As we have argued above, the two different cases have very different implications for action by the centre. In the first case, benchmarking is not prescriptive about the way efficiency should be improved.

Instead, managers are given targets and high-powered incentives to meet them. Those who succeed are rewarded, those who fail are not. This kind of approach has been used for secondary schools, which receive funding based on the number of pupils which they attract, as noted in Chapter 3. In the second, benchmarking is used to encourage the diffusion of innovation and good practice, but high-powered incentive schemes are not necessary. Those who are the innovators gain from the enhancement of their reputations. Those who are less efficient are told *how* to improve.

The differences in the actions that the centre needs to take under the two cases highlight the need for econometric work to identify which case the health service falls close to. The mere observation of differences between producers is not sufficient evidence to conclude that there is wide variation in types of producers. What is needed are comparisons that correct for differences in costs that are beyond the providers' control. The difficulty in such an exercise is deciding what is within and what is outside the providers' control. The CCI measures recently developed in the NHS assume that many sources of cost difference are beyond managers' control. If this view is taken, the variation in managable costs between different hospitals is not large. From this it could be concluded that the problem is not one of great disparity of practice. On the other hand, there seem to be considerable variation in outcomes within similar hospitals. Only when outcomes and costs are put together into one measure will a true picture of the diversity or similarity of performance be available.

4.3 The form of the benchmark

We have noted several problems that arise in the construction of cost benchmarks. More details are provided in Street (1999), and Dawson and Street (1998). Here we focus on three issues that seem to have received less attention in the UK health literature, but have been important in the use of benchmarks in the utilities, particularly the water industry (see Section 3.2, above).

Controlling for exogenous factors

To be a useful indicator of effort, differences in cost and outcome which are beyond a unit's control must be controlled for. A simple way of doing this is to group hospitals into types and to compare within type. This approach has been adopted for the NHS clinical indicators, where indicators are presented for groups of hospitals and groups are defined according to size and type of hospital (teaching, acute etc).

Another approach is to standardise, through statistical analyses, for differences which are judged to be beyond the control of the agents whose performance is being assessed. This is the approach taken in the water industry and is the approach adopted in the CCI indices. The agent is, by this method, compared against the average agent of its type. With sufficient numbers of agents, gaming by the agent to influence the benchmark against which they are to be judged will be limited as the baseline performance is determined by the average behaviour of all agents. But if there are small numbers of agents then gaming becomes feasible and it is also difficult to carry out regression analysis with any precision. The small number of units in the water industry has meant that is has been difficult to identify with any precision the impact of certain factors. This has led to the use of judgement by the regulator, which renders the whole process far less transparent, and opens the door to the regulated companies to influence the regulator. This may be less of a problem in the NHS as the number of units (some 400 Trusts, 100 Health Authorities and 500 PCGs/PCTs) is much larger than in water. However, there is obviously need to check the robustness of indices.

What may be more of a problem in health is defining what are exogenous and what endogenous determinants of costs. Reference costs implicitly treated most determinants as endogenous: the current assumption embodied in the CCI measures is that many more are exogenous. In a world where plant lasts a long time, and where investment decisions are only weakly under the control of managers and clinicians, a view that most determinants of cost (e.g. split site working) are exogenous is probably correct. However, some factors such as

bed numbers and the scale of inpatient activity, currently treated as exogenous in the 3CCI, may be more properly viewed as under the control of management and clinicians. Inclusion of these in the benchmark will therefore hide real differences in efficiency across providers and will make providers seem more similar than they in fact are.

Use of outcome and cost measures

A wide variety of benchmarks have been tried in the NHS. The current drive is to produce benchmarks at specialty level, and to tackle perceived problems with the current cost based performance measures. One problem is that they are all focused on inputs, rather than outcomes. If a system of incentives based on meeting targets for low costs is established, this will have an adverse effect on quality. Treatments with higher short term costs per patient could well be more cost-effective when long term costs and outcomes of care are taken into account. Hence basing the system on short-term financial costs may well be sub-optimal. So the Department of Health is currently expanding the set of benchmarks to encompass measures of output as well as of cost. This is to be encouraged, as a narrow focus on costs has the attendant dangers that all efforts will be focused on cost reduction and that the conveyed mission of the organisation will appear to be about cost reduction alone.

One issue that needs to be addressed is that in general terms, there is a trade-off between cost reduction and quality improvement. Both require effort on the part of clinicians, managers and other hospital staff. (There clearly are some exceptions to this: methods to improve quality may also in some cases reduce costs: an example being measures to reduce death rates which also minimised the costs of being sued for negligence). It is therefore inappropriate to identify best performers in terms of both dimensions simultaneously. If effort put into quality can only be achieved by taking away effort from cost reduction, then the lowest cost performer will not necessarily be that with the highest quality. There will be a tendency for benchmarks to define the best producer as the one which simultaneously achieves lowest cost

and highest quality. But if there is a trade-off between them, getting to this position is not feasible. A similar argument applies in the water industry, where the regulator measures both operating expenditure and capital expenditure. There is a trade-off – operating costs can be made lower by increasing capital expenditure – so it is not feasible for a firm to have the lowest outcomes on both (see Section 3.2).

The danger of too many benchmarks

In general, the complexity of health care probably means a variety of benchmarks is both inevitable and generally helpful: one summary statistic is unlikely to be able to capture the efficiency of a hospital or of an NHS buyer. But this observation brings with it all the issues raised in our discussion of multi-tasking in Section 2.3 above. Unless the ease of measurement of all tasks is equal, some tasks will be easier to benchmark than others, and agents will concentrate their efforts on these, provided there is some positive reward for doing so. This has the danger that the output of the benchmarked organisation will be distorted towards the measured activities. Even if all tasks can be easily measured, agents may simply lose focus and direction. It is probably better to avoid too many targets.

One response might be to have a few benchmarks in operation at each date, and then change these over time. There is a danger however, that frequent changes in performance measures mean that they all come to be ignored or are seen by those 'at the coalface' as evidence that the centre is unclear about its mission and/or doesn't understand what those delivering health care actually do.

A more useful response would be to identify a few key targets for each type of organisation, but not necessarily have the same benchmarks for each organisation. This would allow each organisation, and those within it, to focus on their key business, whilst allowing the NHS as a whole to meet a range of performance standards. The trick is identifying the correct tasks and targets for each organisation. Picking targets which can be only weakly influenced by the effort of those in the organisation is not sensible. While this may seem obvi-

ous, some of the *Health of the Nation* targets set by the previous government for Health Authorities were very difficult for Health Authorities to achieve, as they could do little to change the level of the targeted outcomes.

In terms of choosing targets, an obvious starting place is the current set of high level performance indicators. They are designed to support the drive by the Department of Health to set, deliver and monitor standards. The indicators were explicitly designed to be used for benchmarking the performance of NHS organisations. They cover both throughput and output, so move the focus away from costs to quality and to outcomes. They are limited in number (41 in all). In terms of the criteria for choice of benchmarks, some of the measures – for example teenage pregnancy and suicide rates – are less under the control of an identified organisation than others – for example rates of surgical intervention – and therefore are not an ideal first choice. Several of the measures are (explicitly designed to be) measures of the product of interagency working – for example discharge of elderly patients - but this does make them difficult to use to assess the performance of a single organisation. However, from the 41, plus the clinical indicators that the Department of Health has recently published, it should be possible to identify a small set of indicators, each applicable to different parts of the NHS.

4.4 The role of high-powered incentives

Should benchmark performance be related to explicit financial rewards?

Currently, benchmarks in the NHS are not linked directly to financial rewards. There are indirect links, and all of the uses listed in Section 4.2 above for benchmarks can be seen as having some indirect financial pay-off for the benchmarked organisation. For example, if benchmarks are used to improve practice, being at the top of the league table may enhance the careers of the professionals who provide the services. For Trusts, indirect financial rewards from engaging in best practice

occur if being at the top of the league table translates into higher demand for the services of the Trust or if being at the bottom translates into lower demand. In addition, comparative cost information may be used by purchasers in negotiating contracts if purchasers of health care have leeway to choose providers.

These are all indirect measures. Are they sufficient? There are at least five arguments that could be made against linking benchmarks more directly to financial rewards or penalties. The first is that individuals in the NHS are not motivated by financial concerns, and that tying benchmarks to financial rewards will have no effect. This argument can be dismissed fairly easily. First, while it may be the case that managers and medical professionals have goals other than their current financial rewards, it is also clear that financial rewards do form part of their utility function[7]. There exists a fairly large body of evidence to show that medical professionals respond to financial incentives (a recent example is the response to the GP fundholding scheme which showed that GP fundholders changed their prescribing behaviour and possibly also referrals patterns when becoming budget holders. Second, at the level of the organisation (a Trust or a health purchaser) small amounts of extra money can have a large impact. When a large proportion of costs are fixed, small decreases or increases in income make a large difference to the financial health of the organisation. For example, it is clear that Trusts pursued GP fundholding purchasers, even though the sums of money they brought in were small relative to the total income of the Trust. Where most income is committed, a small increase may have a large effect of the financial well-being of the organisation, and so also on the welfare of the staff in the organisation. As an illustration of the last point, Courty and Marschke (1999) found that public job placement agencies undertook considerable effort to secure bonuses that amounted to less than 10 per cent of the

7 There is considerable general evidence that doctors are partly motivated by financial rewards. For US literature see Hellinger (1996), Pauly (1980). For the UK see Croxson *et al* (1998), Whynes *et al* (1995).

annual budget of the training centre. Thus financial incentives will matter, and in fact matter a lot. Small financial inducements may offer relatively high powered incentives.

Second, it could be argued that financial rewards are not needed: merely publishing the benchmarks will make Trusts and commissioners improve their performance. However, experience suggests this is not the case. The evidence assembled by YHEC (1999) cited above suggests that Trusts may try not to be at the lower end of the distribution, but this may not affect the behaviour of anyone whose performance is above the bottom. Any effect a published benchmark may have will be attenuated if there is a lot of random noise in the benchmark. If this is the case, then the recorded performance of any one Trust will vary from year to year simply because of random noise[8]. The consequence is that a Trust or purchasing organisation that is at the bottom of the league table one year may well not be the next year, without any action on its part. The Norwegian experience of use of benchmarks supports this: hospitals did little about their cost differences until these differences were linked to financial rewards and penalties (Richmond House Workshop, 1998).

Third, one could take the view that publication of benchmarks and identification of poor performers – a 'name and shame regime' – would be sufficient. This kind of regime has developed in education and to some extent in local government (see Chapter 3 above). While it is true that few parts of the NHS or individuals within it would wish to be 'named and shamed', naming and shaming has several drawbacks. The first is the same argument as advanced above: if there is a large stochastic element to the benchmark, performance against the benchmark will have a random component, which will reduce the effectiveness of being named and shamed. The second is that the NHS itself may not want to run a 'name and shame' regime. Such a

8 E.g. a Trust could experience a worse than average flu epidemic causing its recorded mortality rates to rise, or could face an especially tight labour market, causing its costs to rise as it brought in extra agency nurses.

regime would focus attention on poor performers and might lead those working within the organisation to believe that the main aim of the centre was to humiliate individuals, or that the centre saw them as people without professional pride. If the centre has a problem with conveying a positive mission for the organisation, a naming and shaming regime would be the wrong direction to go in.

The fourth argument against linking benchmarks directly to financial rewards or penalties is that in a competitive market benchmark performance, if the benchmark correctly ranks the quality and cost dimensions that the buyer cares about, will be rewarded through sales and profits. In the quasi-market of the NHS, if benchmarks are used by NHS buyers to make commissioning and investment decisions, then Trusts will be rewarded for their good performance through the commissioning process. So there is no need for the centre to link benchmark performance to financial rewards or penalties: the behaviour of NHS buyers will do this. But at present, the incentives offered by the commissioning process alone seem to be weak. Under the internal market arrangement, NHS buyers had, in theory, the ability to shift contracts between providers. Yet commentators such as Dawson and Street (1998, p 5) have argued that even under the operation of the internal market Health Authorities could, or would, not shift their main contracts. Hence there were few rewards to providers for being top of the league table, or few penalties for being at the bottom. Under the current arrangements, the main vehicle for commissioners to exert influence over Trusts is the Long Term Service Agreements. These are still at an early stage of development in many areas. Once developed, they will be in operation for a minimum of three years. They will involve a large number of parties (PCGs/PCTs as lead commissioners, social services, voluntary groups, and other local authority bodies). These arrangements mean that the threat of losing contracts as a consequence of a poor benchmarking performance is small. If purchaser organisations had been in operation for a long period, and there had been a stable set of benchmarks in place throughout that time, then this mechanism might be all that is

required. There would be no need to devise complex financial arrangements to reward well performing providers and penalise bad ones. But it appears that the current commissioning arrangements are unlikely to reward the high performers and penalise the poor very heavily (if at all). The novelty of the current purchasing arrangements are likely to further dilute the impact of published benchmarks on purchasing behaviour[9].

Whatever indirect arrangements are used to reward providers for good performance, this mechanism does not address the link between rewards and good performance for commissioner organisations. If patients could choose between commissioners and if published performance standards were credible, then patients exercising their choice between purchaser would reward the good and penalise the poor quality. This is, at least in theory, what benchmarking in education is intended to achieve. However, currently patients cannot choose their commissioning organisation since patients have little choice of GP. The creation of PCGs/PCTs, which cover a wider geographical area, makes patient choice of purchaser even less possible than under the GP fundholding scheme where at least patients could in theory change purchaser by switching GP.

Fifth, the career concerns of managers, clinicians, and other professionals in the health service might mean that financial rewards are unnecessary. Simply being known to be a good clinician or commissioner may be a strong incentive for a group who have strong professional norms and career concerns. Being identified as a good performer (e.g. beacon status) will be sufficient to induce current and continued effort. This relies on the career concerns of the agents being aligned completely with those of the principals, and there being an active labour market in which good performers can be rewarded

9 Note that this mechanism relies on purchasers using the benchmarks to reward good performance. This in turn means that the benchmarks must convey information relevant to the buyers, which has implications for the design of benchmarks if the purchasing process alone is to be used.

through promotion, either within their own purchaser organisation or by movement to a larger one. However, if the priorities of clinicians do not reflect those of the Trust management, or the PCGs do not reflect those of their patients, rewards based on the priorities of the agents will not be sufficient. The centre may then still wish to introduce links between benchmark performance and financial rewards. In addition, reward through promotion rewards only the individual: it may be desirable to reward a larger group of individuals. Unlike promotion, financial rewards tied to performance against benchmarks can be used to reward the organisation.

Achieving any benchmark measures will not be the sole objective of the organisation. This is undoubtedly true: as we have argued at length, government agencies typically have several, often conflicting, objectives. The NHS is no exception. Its many goals include reducing costs, increasing quality, promoting research and innovation, meeting the needs of patients, as well as perhaps more political requirements of delivering a service which is equitable, which meets the desires of those working in the NHS, and the political ambitions of key Ministers. These goals get translated into multiple tasks. Some of these tasks may be easily measured, others not. High powered incentives linked to benchmarks may therefore result in problems of the kind outlined in Section 2.3 above. In brief, these include focusing effort only on the measured tasks, misrepresenting outputs, and even lowering the total amount of effort. The various forms such activities may take in the NHS, and evidence for their existence has been reviewed in Section 4.1. This argument is in our view the strongest one, and indeed is used to explain why payment by results is not common for complex jobs. However, it is not an argument against linking payment to performance against benchmarks per se. It is one that says that the benchmark chosen must measure something the organisation wishes to promote, even at the expense of other less measurable activities, since there will be a cost in terms of having less measurable goals performed less well.

The form of the benchmark

Given this discussion, there is a case for linking financial rewards to performance against benchmarks. But in that case the literature suggests that the number of benchmarks applied to each group of agents must be limited. While the centre may have many goals, translating each of these into a benchmark for all organisations within the NHS is counterproductive. Having many benchmarks raises all the problems identified in the multi-tasking and career concern literature. More general literature on the behaviour of government agencies (Wilson, 1989) suggests that successful agencies pursue narrow and clear missions. Having many benchmarks creates the impression that the centre doesn't know what its mission is. Each individual agent should face only a limited set of benchmarks. One way of doing this, while at the same time ensuring that the centre meets a range of objectives, is to set a limited number of benchmarks for each type of organisation within the NHS.

Additionally, benchmarks should include quality as well as cost measures. The literature on multi-tasking indicates that if a benchmark is set for one aspect of production, there is a danger that other aspects will be ignored. The development of reference costs has focused upon costs as the important aspect of production. It is clear that costs are not the only dimension of relevance. The Department of Health has recognised this and is now making available measures of quality. Ideally, measures that encompass both cost and quality would be used. In their absence it is important that quality benchmarks are widely promoted, and that their development should receive as much attention as that of costs. This suggests there needs to be rapid movement towards measures of quality adjusted for casemix (e.g. risk adjusted mortality outcomes).

One type of benchmark that could be operated at below single organisation level would be outcome measures for individual consultants or consultant teams. These clinicians are central to the performance of a Trust, and the quality of their output is of vital importance for users and so for commissioners. Enthoven (1999) presents ample

evidence which shows that the publication of risk adjusted mortality outcomes for clinicians in the US has led to quality improvement in the system through both retirement and learning.[10]

Benchmarks should be defined relative to the performance of all groups/organisations whose activities are being measured by the benchmark. If there are sufficient numbers of a similar type, one organisation's behaviour will not significantly affect the value of the benchmark. This will reduce the extent of manipulation of data to alter the benchmark. If, on the other hand, the benchmark for an organisation is defined relative to its own performance (for example, costs or quality achieved last year), this reduces incentives for effort and it will probably encourage misrepresentation to alter the value of the benchmark.

Should the financial rewards be at the level of the organisation or at individual level?

Rewards at the level of the organisation are essentially rewards to a team, where a single team is defined at the level of the whole organisation. There are several arguments for team rewards in health care. First, team rewards may work better because delivering health care is a highly interdependent activity. An individual reward scheme placed in such a context could be arbitrary, costly to operate, and possibly encourage responses that lowered or diverted effort from the production of health care. Second, teams may also be a more efficient way to deliver health care, as they may encourage co-operation and inter-professional working, and they allow monitoring by groups of employees who can have better information than managers, so discouraging shirking and increasing effort.

As noted in Section 2.3, while team rewards may in theory be prone to free-riding, this does not appear to be the case in the private

10 In the US context, publication of risk-adjusted mortality outcomes is not simply 'naming and shaming' as clinicians' rewards are generally linked to the number of patients that they treat.

sector. It might be argued that these results arise because the performance indicator (profit) is not contentious. However, evidence from a large federal job training programme in the US, deriving from the Job Training Partnership Act (JTPA), suggests that team rewards for public sector organisations may also work. The JTPA was set up with a system of explicit incentives for the training agents based at agency level. Under the JTPA system, local training centres receive monetary rewards based on the employment levels and wage rates attained by graduates of the programme. Courty and Marschke (1999) showed that managers responded to these relatively small, team based, incentives by altering the timing of their graduation dates to maximise the chance of getting the team bonus. Managers were postponing the graduation dates of trainees who did not have employment at the end of their training for as long as possible (up to 90 days) in the hope that the trainees would obtain employment, thereby improving the probability of managers earning rewards for favourable training outcomes. Heckman *et al* (1996) investigated whether the teams responded to the incentive to 'cream-skim' the most employable of the applicants into the programme. They estimated the probability of acceptance into the programme using predicted earnings levels and earnings gains, calculated from observed human capital variables, as independent variables. In fact, they find that people with lower expected earnings levels are significantly more likely to be accepted into the programme. Weaker evidence suggests that those with larger expected gains are more likely to be accepted. These results contrast with the cream-skimming prediction, suggesting instead that preferences of the agency workers for helping the disadvantaged override pecuniary incentives in this case.

So it does not seem that group schemes only work for private organisations. One distinguishing feature of the JTPA that is worth noting is that, compared to many government organisations, the training agents have a well-defined set of goals. These are not profit, but are probably relatively uncontentious to the workers in the scheme. So what this literature perhaps suggests is that if rewards are to be defined

at team level, the rewards must be linked to outputs and/or outcomes which all team members view as important.

In addition, payment at team level must deal with the fact that teams in the NHS (Trusts or commissioners) are composed of individuals for whom the relative contribution of current NHS pay to other rewards may be very heterogeneous. Consultants may derive a large part of their income from private practice. Career concerns will be more important to professionals than to manual staff. So if financial rewards are given at team level, they will have to distributed within the team in a way that ensures that those whose effort is central get sufficient rewards from undertaking this effort. This does not mean that rewards should be in strict proportion to NHS pay. It may be the case that career concerns of certain agents give them motivation which is above that of those without career concerns, and so mean the former group require less from the incentive scheme relative to their pay.

The definition of a team is not an easy issue and depends on the interdependence of tasks. Teams could be defined at below Trust or buyer organisation level if different teams are responsible for delivering different and discrete outputs. Alternatively, if there is a lot of jointness in the output of teams, defining teams at below Trust level may be difficult. It could also be counterproductive if it reduces co-operation between units that need to work together.

One set of teams that could be defined within Trusts would be consultant-led teams. Different benchmarks could be defined for each consultant team. To avoid competition between teams within a single provider, each team's performance would be compared against the national average. Teams could share rewards as they internally decided (perhaps by a mixture of simple sharing and sharing based on NHS pay). The advantages of consultant based teams are: first, that consultants are pivotal to the output of a hospital so incentives at this level are probably desirable; second, provided adjustment were made for casemix, consultants' costs or outcomes should be comparable; and third, some of the multi-tasking issues that would arise at Trust level could be avoided. There would still remain a multi-tasking issue at the

consultant level, but as we have pointed out, multi-tasking is likely always to be an issue in health care. The disadvantage with establishing consultant led- teams is that outdated methods of working could become entrenched, and that it might increase the extent of competition for resources within Trusts, which could have a detrimental effect on output and morale. Making financial rewards dependent on the performance of other teams within the same Trust might be one way of overcoming this potentially destructive competition. Limiting the extent of high-poweredness of the rewards would be another way of achieving the same end.

5 CONCLUSION

A series of recommendations for the use of benchmarking in the NHS follow from the analysis set out in this paper.

Recent moves by the Department of Health to speed the pace of development of benchmarking are in the right direction. The development of reference costs has focused upon costs as the important aspect of production but it is clear that costs are not the only dimension of relevance. The Department of Health has recognised this and is now making available measures of quality. Ideally, measures that encompass both cost and quality would be used. In their absence it is important that the development of quality benchmarks receives as much resource and attention as that of cost benchmarks, and that their use is then promoted. There needs to be rapid movement towards measures of quality adjusted for casemix (e.g. risk adjusted mortality outcomes). Failure to do this would convey the signal that the NHS is primarily interested in providing a cheap service rather than providing a quality service at value for money. An emphasis on cost will lower the spirit of the workforce and affect recruitment. Such a change in perceived mission may take many years to reverse.

It appears that there is considerable scope to link performance against these benchmarks to financial incentives, as small amounts of money can have a large impact on organisations which have annual budget constraints. So there is scope for the introduction of high-powered incentives in the NHS. But we think that in most areas we would not expect to see, even in the very long run, as high-powered a system as arises in the utilities.

The general asymmetry of information is probably no greater a problem in the NHS than in the utilities, where the inability of the regulator to accurately observe what can be achieved is a well-established problem. But the asymmetry of information between the NHS and individual trusts may affect the speed of development of benchmarks and their links to financial incentives. The utilities have shown that it takes a considerable amount of time to develop information on companies' performance and that it is almost impossible to develop a fully accurate model. In the utility context shareholders' wealth pre-

sents a buffer between errors and performance. If the benchmark is set too tightly for a company its shareholders rather than consumers are likely to suffer first. There is no analogue in the NHS where there is little flexibility in budgets and there is no surplus as such. It follows that small changes in overall budget numbers can have large effects. This may be good for the operation of incentives but does suggest that initially, if the incentive regime is extremely high-powered, there may have to be a softening of the consequences of failure to meet the target or acceptance of a short term fall in standards. The former unravels the incentive mechanism while the latter defeats the purpose of the approach.

It may be possible to detatch incentives from expenditure on patient care. This could be accomplished by creating an incentive fund which was 'top-sliced' from budgets. For example, the overall budget for GP remuneration could be top-sliced and given to PCGs/PCTs to fund incentive schemes. This would mean that, if a particular GP failed to meet prescribing quality targets, her patients would not suffer further because her prescription budget is reduced. Such a scheme would also require that the incentive went to the GP as income, not as a fund for spending to the benefit of practice patients. The result would be that quality of care would improve even though there was no change in average GP remuneration and no change in expenditure on patients.

It follows that benchmarking with attendant financial incentives should be initially introduced with not overly high-powered incentives. These incentives can then be ratcheted up over time. The speed at which this should occur will depend on the ability of the statistical modelling to pick up the genuine differences between decision makers (either organisations such as Trusts and buyers, or agents within these organisations) that cannot be changed in the medium term. The identification of fixed effects in NHS benchmarks is extremely important and requires more attention. Further statistical modelling is needed to isolate the impact of factors that are under the control of decision makers on cost and quality.

The greatest difficulty for the use of high-powered incentives arises because of the multi-tasking aspects of so many of the organisations in the NHS. No single benchmark measure can reflect the objectives of the organisation. The NHS has many goals, as do the organisations within it. Some of these tasks may be easily measured, others may not be. We have discussed earlier the effect of multi-tasking where there is asymmetry in the ability of measuring tasks. In particular, there is a distortion of effort away from the tasks that are harder to measure. This is a far larger problem in the NHS than in utilities or most situations where benchmarking is used. The effect is that the incentive structure cannot be pushed strongly unless there is a reduction in the value attached to the less measurable tasks. We do not consider that such a change has taken place (or indeed that it necessarily will) and so suggest that incentives must inevitably be less high-powered than elsewhere. To ignore this fact and to push 'too hard' may well have genuinely damaging side effects.

The presence of multi-tasking and multiple principals also appears to be the biggest factor that will determine which activities are to be linked to financial rewards and how strong the incentives should be. Since the extent of asymmetry of measurement of multi-tasks will itself differ significantly between activities this suggest that there are strong arguments for focused benchmarks and incentive systems. Activities in a well-defined small organisation (say a pathology laboratory) may suffer less from the multi-tasking problem. The broader the activity measured the bigger the multi-task problem becomes for the group of individuals who undertake these activities. The ability to incentivise individuals to expend effort on a narrow range of services (e.g. the production of services from pathology laboratories) will be lost if these services are bundled together with activities that have difficult multi-tasking problems. The difficult-to-measure tasks in other activities will be damaged if high powered incentives are applied to all measurable tasks within the grouping. This suggests that it is useful to identify activities at a micro level and to use higher powered incentives where there is little asymmetry of measurement within that activity.

There will still be restrictions on the amount of financial incentives that can be used for these smaller activities, since an organisation faced with strong financial incentives for some activities and weaker financial rewards in others will have an incentive to transfer effort to those activities which are rewarded (and are easier to measure). That is, there is a multi-tasking problem at the organisation level as well as the lower activity level. The ring-fencing of incentive schemes to particular activities will mean that all tasks in the weaker-to-measure category will suffer (equally) if too high incentives are focused on activities where there is no asymmetry problem. The net effect is that drilling down to activities is useful but does not solve the whole problem.

The number of benchmarks faced by any part of the NHS needs to be limited. The NHS has many goals. It is a government agency with several principals all of whom have different objectives. While the centre has many legitimate goals, translating each of these into a benchmark for all organisations within the NHS is counterproductive. The use of many benchmarks for each organisation creates the impression that the centre doesn't know what its mission is. This is at best confusing, but worse, may lead to individuals reducing total effort. Uncertainty over the nature of tasks to be carried out by each worker or the effort allocation between each task also reduces effort. Successful agencies appear to pursue narrow and clear missions. In benchmarking terms this suggests a limited number of benchmarks for each type of organisation within the NHS should be set, but with different benchmarks set for different organisations.

REFERENCES

Adnett, Nick, and Davies, Peter (1999), 'Schooling Quasi-Markets: Reconciling Economic and Sociological Analyses', *British Journal of Educational Studies*, 47/3, pp 221-234.

Ahmed, Pervaiz K, and Rafiq, Mohammed (1998), 'Integrated Benchmarking: A Holistic Examination of Select Techniques for Benchmarking Analysis', *Benchmarking for Quality Management and Technology*, 5/3, pp 225-242.

Auriol, Emmanuelle, Pechlivanos, Lambros and Friebel, Guido (1999), 'Teamwork Management in an Era of Diminishing Commitment', CEPR Discussion Paper 2281.

Ball, Amanda, Bowerman, Mary and Hawksworth, Shirley (1999a) 'Great Expectations: Benchmarking for Best Value', Unpublished Paper, Sheffield University Management School.

Ball, Amanda, Bowerman, Mary and Hawksworth, Shirley (1999b), 'Benchmarking in Local Government under a Central Government Agenda', Unpublished Paper, Sheffield University Management School.

Bartlett, Will, and Le Grand, Julian (1993), 'The Theory of Quasi-Markets', in Le Grand and Bartlett (eds), *Quasi-Markets and Social Policy*, Macmillan, Basingstoke.

Bovaird, Tony (1999), *Achieving Best Value through Competition, Benchmarking and Performance Networks*, Warwick/DETR Best Value Series, Paper 6.

Bovaird, Tony, and Davis, Paul (1999), 'Learning to Manage within Limited Resources: Coping Strategies and Learning Breakthroughs in UK Local Government', *International Journal of Public Sector Management*, 12/3, pp 293-313.

Boyne, George (1997), 'Comparing the Performance of Local Authorities: An Evaluation of the Audit Commission Indicators', *Local Government Studies*, 23/4, pp 17-43.

Boyne, George, Gould-Williams, Julian, Law, Jennifer, Marriott, Neil and Walker, Richard (1999), *Wales Evaluation Study on Best Value*, Cardiff Business School Public Services Research Unit, Working Paper 2.

Bradley, Steve, Johnes, Geraint and Millington, Jim (1999), 'School Choice, Competition and the Efficiency of Secondary Schools in England', CREE Discussion Paper, Department of Economics, Lancaster University.

Bullivant, John (1996), 'Benchmarking in the UK National Health Service', *International Journal of Health Care Quality Assurance*, 9/2, pp 9-14.

Bullivant, John (1998), *Benchmarking for Best Value in the NHS*, FT Healthcare, London.

Burns, William (ed) (1992), *Performance Measurement Evaluation and Incentive*, Harvard Business School Press.

Cabinet Office (1999), *1998 Next Steps Report*, Cm 4273, Cabinet Office, London.

Coopers & Lybrand (1994), *Survey of Benchmarking in the UK*, Coopers & Lybrand and CBI National Manufacturing Council, London.

Courty, Pascal (1997), 'Strategy Communication and Measurement Systems', Unpublished Paper, Department of Economics, Universitat Pompeu Fabra.

Courty, Pascal, and Marschke, Gerald (1999), 'An Empirical Investigation of Gaming Responses to Performance Incentives', Unpublished Paper, London Business School.

Cowper, Jeremy, and Samuels, Martin (1997), *Performance Benchmarking in the Public Sector: The United Kingdom Experience*, Cabinet Office, London.

Croxson, Bronwyn, Propper, Carol and Perkins, A (1998), 'Do Doctors Respond to Financial Incentives? UK Family Doctors and the GP Fundholder Scheme', CMPO Working Paper 98/001, University of Bristol.

Cubbin, John, and Tzanidakis, George (1998), 'Regression versus Data Envelopment Analysis for Efficiency Measurement: An Application to the England and Wales Regulated Water Industry', *Utilities Policy*, 7, pp 75-85.

Davies, AJ, and Kochhar, AK (1999), 'Why British Companies Don't Do Effective Benchmarking', *Integrated Manufacturing Systems*, 10/1, pp 26-32.

Davis, Howard, and Walker, Bruce (1998), 'Contracting and Best Value: Developing a New Approach', *Local Governance*, 24/2, pp 111-118.

Davis, Paul (1998), 'The Burgeoning of Benchmarking in British Local Government: The Value of 'Learning by Looking' in the Public Services', *Benchmarking for Quality Management and Technology*, 5/4, pp 260-270.

Dawson, Diane, and Street, Andrew (1998), 'Reference Costs and the Pursuit of Efficiency in the 'New' NHS', University of York, Centre for Health Economics Discussion Paper 161.

Dence, Roger (1995), 'Best Practices Benchmarking', in Jacky Holloway, Jenny Lewis and Geoff Mallory (eds), *Performance Measurement and Evaluation*, Sage Publications, London.

DETR (Department of the Environment, Transport and the Regions) (1998), *Modern Local Government: In Touch with the People*, Cm 4014, The Stationery Office, London.

DfEE (Department for Education and Employment) (1997), *Excellence in Schools*, Cm 3681, The Stationery Office, London.

DfEE (Department for Education and Employment) (1998), 'Value Added Pilot', http://www.dfee.gov.uk/ performance/vap_98/

DfEE (Department for Education and Employment) (1999), 'Value Added in the 16-18 Performance Tables', http://www.dfee.gov.uk/valueadded

Dixit (1997), 'Power of Incentives in Public versus Private Organisations', *American Economic Review Papers and Proceedings*, 87:2, 378-382.

Drago, R, and Heywood, J (1995), 'The Choice of Payment Schemes: Australian Establishment Data', *Industrial Relations*, 34, pp 507-531.

Drake, Leigh, and Weyman-Jones, Thomas G (1996), 'Productive and Allocative Inefficiencies in UK Building Societies: A Comparison of Non-Parametric and Stochastic Frontier Techniques', *Manchester School*, 64/1, pp 22-37.

Drew, Stephen AW (1997), 'From Knowledge to Action: the Impact of Benchmarking on Organizational Performance', *Long Range Planning*, 30/3, pp 427-441.

Eccles, Robert (1991), 'The Performance Measurement Manifesto', *Harvard Business Review*, January/February, 1991, pp 131-137.

Enthoven, A (1999). Rock Carling Lecture. Nuffield Trust.

Europe Economics and Professor Nick Crafts (1998), *Water and Sewerage Industries: General Efficiency and Scope for Improvement, Final Report for Ofwat*. Europe Economics, London.

Evans, Jennifer and Vincent, Carol (1997), 'Parental Choice and Special Education', in Ron Glatter, Philip A Woods and Carl Bagley (eds), *Choice and Diversity in Schooling*, Routledge, London.

Gewirtz, Sharon, Ball, Stephen J and Bowe, Richard (1995), *Markets, Choice and Equity in Education*, Open University Press, Buckingham.

Glennerster, Howard, and Hills, John (eds) (1998), *The State of Welfare: The Economics of Social Spending*. Oxford University Press.

Goddard, Maria, Mannion, Russell and Smith, Peter C (1998), 'The NHS Performance Framework: Taking Account of Economic Behaviour', University of York, Centre for Health Economics Discussion Paper 158.

Goldstein, Harvey and Thomas, Sally (1996), 'Using Examination Results as Indicators of School and College Performance', *Journal of the Royal Statistical Society,* A 159, pp 149-163.

Gordon, Liz, and Whitty, Geoff (1997), 'Giving the 'Hidden Hand' a Helping Hand? The Rhetoric and Reality of Neoliberal Education Reform in England and New Zealand', *Comparative Education,* 33/3, pp 453-467.

Hardman, Jason, and Levacic, Rosalind (1997), 'The Impact of Competition on Secondary Schools', in Ron Glatter, Philip A Woods and Carl Bagley (eds), *Choice and Diversity in Schooling*, Routledge, London.

Heckman, JJ, Smith, JA and Taber, C (1996), 'What Do Bureaucrats Do? The Effect of Performance Standards and Bureaucratic Preferences on Acceptance into the JTPA Program', NBER Working Paper, 5535.

Hellinger, F (1996), 'The Impact of Financial Incentives on Physician Behaviour in Managed Care Plans: A Review of the Evidence', *Medical Care Research Review*, 53, pp 294-314.

Higgs, G, Bellin, W, Farrell, S, and White (1997), 'Educational Attainment and Social Disadvantage: Contextualizing School League Tables', *Regional Studies,* 31/8, pp 775-789.

Holloway, Jacky, Francis, Graham, Hinton, Matthew and Mayle, David (1998), 'Best Practice Benchmarking: Delivering the Goods?' *Total Quality Management,* 9, 4-5, pp S121-S125.

Holloway, Jacky, Hinton, Matthew, Francis, Graham and Mayle, David (1999), *Identifying Best Practice in Benchmarking,* Chartered Institute of Management Accountants, London.

Holmstrom, Bengt (1979), 'Moral Hazard and Observability', *Bell Journal of Economics* 10, pp 74-91.

Holmstrom, Bengt, and Milgrom, Paul (1991), 'Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership and Job Design', *Journal of Law, Economics and Organisation,* 7, pp 24-52.

Kandel, E, and Lazear, E (1992), 'Peer Pressure and Partnerships', *Journal of Political Economy,* 100/4, pp 801-817.

Kaplan, Robert, and Norton, David (1996), *The Balanced Scorecard,* Harvard Business School Press, Boston, Massachusetts.

Locke, Edwin, and Latham, Gary (1996), *A Theory of Goal Setting and Task Performance,* Prentice-Hall, Englewood Cliffs, NJ.

MMC (Monopolies and Mergers Commission) (1996) *Severn Trent plc and South West Water plc,* HMSO, London.

MMC (Monopolies and Mergers Commission) (1997), *Report on Northern Ireland Electricity,* The Stationery Office, London.

Modernising Government (1999), Cm 4310, The Stationery Office, London.

Monkhouse, Elaine (1995), 'The Role of Competitive Benchmarking in Small to Medium-Sized Enterprises', *Benchmarking for Quality Management and Technology,* 2/4, pp 41-50.

National Audit Office (1998), *Benefits Agency: Performance Measurement.* HC 952, 1997/98, The Stationery Office, London.

Nutley, S, and Smith, Peter (1998), 'League Tables for Performance Improvement in Health Care', *Journal of Health Services Research and Policy,* 3/1, pp 50-57.

Offer (1999), *Review of Public Electricity Suppliers, Distribution Price Control Review: Consultation Paper,* Offer, Birmingham.

Ofwat (1994), *Future Charges for Water and Sewerage Services: The Outcome of the Periodic Review,* Ofwat, Birmingham.

Ofwat (1998a), *A Benchmarking Study of the England and Wales Water Companies and Sydney Water Corporation Ltd for 1996-97,* Ofwat, Birmingham.

Ofwat (1998b), *Assessing the Scope for Future Improvements in Water Company Efficiency: A Technical Paper,* Ofwat, Birmingham.

Ofwat (1999a), *Comparing the Performance of England & Wales Water & Sewerage Companies with Sydney Water and Water Corporation Western Australia, 1997-98 data,* Ofwat, Birmingham.

Ofwat (1999b), *Draft Determinations: Future Water and Sewerage Charges 2000-05,* Ofwat, Birmingham.

PA Consulting Group (1999), *Final Report for Ofwat on Operational Process Benchmarking and Cost Reduction,* PA Consulting Group, London.

Partnership Sourcing (1997), *Benchmarking the Supply Chain: First Cycle of Surveys,* Partnership Sourcing.

Pauly, MV (1980), *Doctors and their Workshops,* NBER Monograph, University of Chicago Press, Chicago.

Power, Sally, Halpin, David and Whitty, Geoff (1997), 'Managing the State and the Market: New Education Management in Five Countries', *British Journal of Educational Studies,* 45/4, pp 342-362.

Richmond House Workshop (1998), 'The Use of Benchmarking 'Reference Costs' to Improve NHS Provider Performance'. Office of Health Economics and University of York Centre for Health Economics.

Samuels, Martin (1998), *Towards Best Practice: An Evaluation of the First Two Years of the Public Sector Benchmarking Project 1996-98,* Cabinet Office, London.

Saunders, Lesley (1999), 'A Brief History of Educational 'Value Added': How Did We Get to Where We Are?', *School Effectiveness and School Improvement,* 10/2, pp 233-256.

Sawkins, John W (1995), 'Yardstick Competition in the English and Welsh Water Industry: Fiction or Reality', *Utilities Policy,* 5/1, pp 27-36.

Shleifer, A (1985), 'A Theory of Yardstick Competition', *Rand Journal of Economics,* 16, pp 319-327.

Smith, Peter (1990), 'The Use of Performance Indicators in the Public Sector', *Journal of the Royal Statistical Society,* Series A, 153, Part 1, pp 53-72.

Smith, Peter (1995), 'On the Unintended Consequences of Publishing Performance Data in the Public Sector', *International Journal of Public Administration,* 18, 2-3, pp 277-310.

Stephens, Amanda, and Bowerman, Mary (1997), 'Benchmarking for Best Value in Local Authorities', *Management Accounting,* 75/10, pp 76-7.

Street, Andrew (1999), 'Interpreting the NHS Cost Indices for Acute Trusts', University of York, Centre for Health Economics Discussion Paper 175.

Taylor, Jim, and Bradley, Steve (1998), 'Cost Functions in Secondary Schools', CREE Discussion Paper 2/98, Lancaster University Management School.

Thomas, Sally, Sammons, Pam, Mortimore, Peter and Smees, Rebecca (1997), 'Stability and Consistency in Secondary Schools' Effects on Students' GCSE Outcomes over Three Years', *School Effectiveness and School Improvement,* 8/2, pp 169-197.

Tirole, Jean (1994), 'The Internal Organisation of Government', *Oxford Economic Papers,* 46/1, pp 1-29.

Tucker, F, Zivan, S and Camp, R (1987), 'How to Measure Yourself Against the Best', *Harvard Business Review,* January/February 1987, pp 8-10.

Voss, Christopher, Par Ahlstrom, A and Blackmon, Kate (1997), 'Benchmarking and Operational Performance: Some Empirical Results', *International Journal of Operations and Production Management,* 17/10, pp 1046-1058.

Whymark, John (1998), 'Benchmarking and Credit Risk Management in Financial Services', *Benchmarking for Quality Management and Technology,* 5/2, pp 126-137.

Whynes, D, Baines, DL and Tolley, KH (1995), 'GP Fundholding and the Costs of Prescribing', *Journal of Public Health Medicine,* 17, pp 323-329.

Williamson, OE (1985), *The Economic Institutions of Capitalism, Firms, Markets and Relational Contracting,* Free Press, New York.

Wilson, James (1989), *Bureaucracy: What Government Agencies Do and Why They Do It,* Basic Books, New York.

Woods, Philip A, Bagley, Carl and Glatter, Ron (1998), *School Choice and Competition: Markets in the Public Interest.* Routledge, London.

YHEC (York Health Economics Consortium) (1999), *NHS Trust and Specialty Unit Cost Benchmarking in the NHS: Evaluation of New Unit Cost Indices for NHS Trusts,* YHEC, York.

RECENT OHE PUBLICATIONS

Primary Care and the NHS Reforms: a Manager's View
by Robert Royce, 2000 (price £10.00)

Narrowing the Gap between Provision and Need for Medicines in Developing Countries
by Hannah Kettler, 2000 (price £7.50)

Risk Adjusting Health Care Resource Allocations – Theory and Practice in the UK, The Netherlands and Germany
by Adam Oliver, 1999 (price £7.50)

Genomics, Healthcare and Public Policy
eds. Paul Williams and Sarah Clow, 1999 (price £10.00)

Doctors, Economics and Clinical Practice Guidelines: Can they be Brought Together?
by David Eddy, 1999 (price £5.00)

Leadership, Change and Primary Care Groups
ed. Louise Locock, 1999 (price £5.00)

Public Involvement in Priority Setting
ed. Lisa Gold, 1999 (price £5.00)

Risk and Return in the Pharmaceutical Industry
eds. Jon Sussex and Nick Marchant, 1999 (price £10.00)

The New NHS: What Can We Learn from Managed Care in New Zealand and the US?
ed. Nick Goodwin, 1999 (price £5.00)

Trade Mark Legislation and the Pharmaceutical Industry
by Shelley Lane with Jeremy Phillips, 1999 (price £10.00)

Organisational Costs in the New NHS
by Bronwyn Croxson, 1999 (price £7.50)

Disease Management, the NHS and the Pharmaceutical Industry
by Anne Mason, Adrian Towse and Mike Drummond, 1999 (price £7.50)

Updating the Cost of a New Chemical Entity
by Hannah Kettler, 1999 (price £7.50)

Controlling NHS Expenditure: the Impact of Labour's NHS White Papers
by Jon Sussex, 1998 (price £5.00)

GP Commissioning Groups – the Nottingham Experience
by Stephen Earwicker, 1998 (price £5.00)

Our Certain Fate: Rationing in Health Care
by Alan Maynard and Karen Bloor, 1998 (price £5.00)