Large Scale Models and Large Scale Problems:

The Case of the Health Services

by Peter Smith

DISCUSSION PAPER 110

LARGE SCALE MODELS AND LARGE SCALE PROBLEMS:
THE CASE OF THE HEALTH SERVICES

Peter Smith

August 1993

The Author

Peter Smith is Senior Lecturer in Economics, Finance and Accountancy in the Department of Economics and Related Studies, University of York.

Acknowledgements

Further Copies

Further copies of this document are available (at price £4.50 to cover the cost of publication, postage and packing) from:

> The Publications Secretary
> Centre for Health Economics
> University of York
> York, YO1 5DD

Please make cheques payable to the University of York. Details of other papers can be obtained from the same address, or telephone York (0904) 433648 or 433666.

ABSTRACT

The delivery of health services represents one of the most daunting challenges confronting contemporary management. The complexity of the issues involved suggests that health service policy makers seeking to allocate scarce resources would find the opportunities offered by large scale modelling attractive. However, the application of such models to the health services has not been a success in the United Kingdom. This paper seeks to explain the failure by describing three important initiatives in the UK National Health Service that have occurred over a period of twenty years. The first is the deployment of the large scale "balance of care" model, which was developed in the 1970s, and sought to allocate local government and health service resources between competing claims. The second is the performance indicator initiative of the 1980s, which concentrated on the measurement of a large number of processes and outcomes in the health sector. The third is the "internal market" reform of the 1990s. These developments were informed, respectively, by the disciplines of operational research, accountancy and economics. Alternatively, they can be thought of as representing the planning, bureaucratic and market views of management. The paper argues that the traditional OR approach fails not because its model is an inadequate representation of reality, but because it does not acknowledge the priorities of the manager or politician who must take responsibility for implementing the model's findings. The overriding concerns of the accountable person are likely to be securing control and avoiding blame, rather than the pursuit of either allocative or managerial efficiency. In this respect, the models of accountants and economists are more attractive to managers and politicians.

1. INTRODUCTION

Large organizations pose enormous problems of managerial control. The United Kingdom National Health Service (NHS) is certainly large, spending 5.5 % of gross domestic product in 1991/92 (Treasury, 1992). Moreover, it seeks to deliver a uniquely complex range of services, using diverse technologies which are very imperfectly understood, to a clientele which is heterogeneous in both its needs and expectations. It is therefore hardly surprising to find that, in seeking to secure control of the resources for which they are responsible, the management of the NHS have from time to time developed a range of quantitative models.

Models might be considered "large scale" for one of two reasons. First, the problem being addressed might be large in scope and implications. Thus, by this yardstick, the model developed by the Resource Allocation Working Party (1976) to distribute NHS finances to geographical areas is unambiguously large scale. However, the model itself is relatively simple in both its structure and data requirements. On the other hand, models might be considered large scale because of the magnitude and complexity of their mathematical formulation. By this criterion, some complex models of medical processes might be considered large (Davies and Davies, 1987). However, although the problem being addressed may be of considerable clinical importance, such models are clearly not large scale in the sense of having major resource implications.

This paper focuses on the use of quantitative models of considerable complexity applied to strategic issues of resource allocation in the NHS. The large number of models - many of which are immensely complex - applied to operational issues such as clinical management and tactical resource allocation are not considered here. The emphasis is on the application of extensive models to problems which are large in scope, and sidesteps the question of what is meant by a large scale model. By either of the criteria discussed above, such models are large scale.

The principal aim of such models is to secure strategic control of NHS resources, and must perforce address deeply political issues. It is therefore important to have a clear picture of the pattern of accountability and control within the NHS. The organization has a highly

1

centralized organizational structure. In England, most hospital services are purchased by the 191 District Health Authorities (DHAs). Each DHA is responsible to one of the 14 Regional Health Authorities, which are in turn responsible to the central NHS Management Executive. Democratic accountability is secured by the ultimate responsibility of a member of the government - the Secretary of State for Health and Social Services - for the performance of the NHS. However citizens have very little means of influencing the nature of services in their immediate locality. Members of District Health Authorities are appointed, not elected, and Community Health Councils, responsible for representing the local interests of patients, in practice have very few powers and little influence.

Organizational control in the NHS has two principal components: resource allocation and performance audit. These serve the role, respectively, of prospective, forward looking control, in which power is delegated from the centre to the periphery; and retrospective, evaluative control, whereby the periphery is held accountable by the centre for its use of resources. Resource allocation from the NHS Management Executive to the regions is by formula, based principally on population (Department of Health, 1989). Regional Health Authorities use similar means to distribute resources to districts. Audit of the use of resources is undertaken by a system of performance review. At the highest level, the Chief Executive of the NHS is accountable to Parliament. The network of accountability then flows down through the regions and districts to individual managers. A range of performance indicators is used to inform the review process (see Section 2.2).

Thus the management of the NHS face problems which would not be unfamiliar to students of the late Soviet Union. Policy is set centrally, and interpreted at the local level. However, local management teams have no systematic or coherent means of inferring the needs and preferences of the local population. DHAs would (presumably) wish to ensure that they allocate the resources at their disposal in a manner which satisfies both the Management Executive and the local population. For its part, the centre wishes to ensure that resources are allocated in accordance with its preferences, and are used efficiently.

The organizational structure of the NHS has changed from time to time over the years, although the basic highly centralized design has remained intact throughout its lifetime. Even

2

recent reforms described more fully in Section 2.3 do not challenge the principles of accountability adumbrated above. The principal innovation is that, instead of providing services directly, DHAs purchase services from a range of providers. The managers of purchased services are then held to account through the terms of formal contracts rather than as hitherto through the internal review process.

The purpose of this paper is to assess the role of large scale models in securing control within a large centralized organization such as the NHS. In order to focus the discussion, the next section describes three notable developments that have occurred over a twenty year period. The first is the large scale "balance of care" model, which was developed in the 1970s, and sought to allocate local government and health service resources between competing claims. The second is the performance indicator initiative of the 1980s, which concentrated on the measurement of a large number of inputs, processes and outcomes in the health sector. The third is the "internal market" reform of the 1990s. There then follows an assessment of the impact that these initiatives had on the management of the NHS. And the final section draws some conclusions concerning the role of large scale modelling in the health sector.

2.   THREE INITIATIVES

2.1   The balance of care model

During the 1970s there was a brief efflorescence of quantitative modelling in the health sector, brought about by rapidly increasing computing power made available at that time. At first the emphasis was on tactical issues. However, by the end of the decade, Boldy and Clayden (1979) were able to detect a distinct shift in emphasis in OR projects from the tactical to the strategic. It is noteworthy that, in response to Boldy and Clayden's survey, most researchers described their methodology as "applied common sense" rather than reflecting traditional OR modelling procedures. Nevertheless, as Gibbs (1981) shows, there were a number of intricate strategic models in existence that went beyond such modest ambitions.

3

Amongst the most celebrated of the large scale health care models of the 1970s was the balance of care model, developed by the Operational Research Service of the central government Department of Health and Social Security (DHSS). The technical formulation of this model is given by McDonald *et al* (1974) and Coverdale and Negrine (1978), and the context within which it was developed by Gibbs (1978). The purpose of the model was to secure a rational allocation of limited resources in the health and personal social services to the competing demands of numerous client groups. Clearly, whether applied at the national or local level, such a model was of strategic importance, and its mathematical formulation and data demands suggest that it was unambiguously large scale. Moreover, the project was doubly ambitious because it sought to straddle organizational boundaries. In the UK, health services are delivered by the NHS, while personal social services are delivered by local government, which has a very different organizational structure and culture to the NHS.

Several variants of the model exist, and space precludes a full mathematical exposition. On the demand side, a number of client groups are identified (six in the example given by Boldy *et al*). Each client group is divided up into a number of reasonably homogeneous categories, of which there are about 150 in total. For each category there usually exist a number of acceptable modes of care, such as mixtures of hospital care, domiciliary care and day facilities. There might be many modes of care for some categories of person. The model links demand with supply by assuming that each mode of care has certain implications for resource use, assuming an ideal level of provision for persons in the category under consideration. Finally, the resource use has financial implications for the organizations responsible for delivering the services - in this case, the NHS and local government.

Providing that all the data requirements are met, this structure is amenable to mathematical programming, and one use of the model is indeed to estimate the "minimum cost" of delivering ideal levels of service to all clients. In this guise the model requires the user to specify the numbers of clients in category i (D_i), the ideal allocation of service k to client i in mode l of care (U_{ilk}), the maximum and minimum availability of resource k (R_k and R_k^*), and the unit cost of resource k (C_k). Decision variables are the number of clients in category i to use mode l (d_{il}) and the amount of service k allocated to client group j (r_{jk}),

where a client group is an aggregation of categories. The cost minimizing linear programme is then

$$Minimize \sum_j \sum_k C_k r_{jk}$$
$$subject\ to$$
$$r_{jk} = \sum_{i \in j} U_{uk} d_{ui}$$
$$\sum_l d_{ui} = D_i \qquad \forall i$$
$$R_k^* \le \sum_j r_{jk} \le R_k \qquad \forall k.$$

The mathematical programme searches for the allocation of clients to resources which minimizes costs, subject to the constraint that every client receives the ideal level of service, and subject to physical resource constraints.

Because of the apparently insatiable demand for health and personal social services, and the impossibility of offering ideal levels of care (however defined) within realistic resource constraints, this formulation of the balance of care model was never intended as a planning device. Instead, it appears to have been envisaged as providing a benchmark against which the results from more realistic formulations might be compared.

In order to make the model operational, a number of new concepts were introduced which acknowledge the scarcity of resources. The variable s_{jk} is defined as the amount of service k allocated to client group j as a proportion of the ideal service level. Then the function

$$h_{jk}(s_{jk}) = A_{jk} s_{jk}^{\phi_{jk}} + B_{jk}$$

is (in some loosely defined sense) a measure of the "worth" of securing a certain level of care for client group j. The parameters A_{jk}, B_{jk} and Φ_{jk} are calibrated from observed behaviour (as a result of which, this formulation is sometimes referred to as the "inferred worth" model).

5

The objective function then reflects the difference between benefits (as reflected in the measures of worth) and costs, and is written

$$Maximize \sum_{j} \sum_{k} h_{jk}(s_{jk}) - C_k r_{jk}$$

The constraints are augmented to acknowledge the total revenue available (Boldy *et al*, 1981). Clearly other formulations along these lines are possible, and details are given by Coverdale and Negrine (1978). However, the principle underlying them all is clear. The models seek to allocate scarce resources between competing claims, and in order to do so require value judgements about the worth of securing certain levels of services for various types of client.

The model was initially developed for use at the national level. However, Gibbs (1978) advocates the use of the model by both the central authority (in setting policy) and at the local level (in developing local plans), and Boldy *et al* argue that the closer links between health and personal social services make it more relevant at the local level. Certainly, providing one accepts the notion of maximizing benefits in relation to costs, the model offers a close representation of the technical problems implicitly facing local health authorities and local governments. Given the minimal costs charged directly to users of their services, demand will always exceed supply, so decision-makers need some means of reconciling competing claims from diverse sources. In seeking to guide decision-makers in this crucial area, the inferred worth model incorporates many intuitively attractive features. It is comprehensive, including all client groups and resources of relevance. In the functional form of the worth functions h(.) it acknowledges the diminishing marginal worth of pumping more resources into a particular client group. It is flexible, in allowing a wide variety of assumptions to be tested, and acknowledges that a client might use several services in a variety of ways. And it tackles in a systematic way an issue of great sensitivity: the interface between the NHS and local government. In short, the model represents the apotheosis of the planning culture that pervaded the NHS in the 1970s.

However, the architects and (presumably) advocates of the model hint at some

6

difficulties that arose in trying to make the model operational. Coverdale and Negrine (1978) admit that "the calibration of parameters has proved to be difficult" and that "establishing the priorities behind services has presented problems". Boldy *et al* (1981) found that, in an application of the model, "when several client groups are considered, it is not easy for managers to comprehend the complex interactions that may take place" and that "the complexity and size of the current model preclude it from being run on a local computer by local managers". Gibbs (1978) seeks to answer some of the obvious criticisms, such as the use of surrogates for the true outputs of the services, and the difficulty of identifying ideal levels of service. However, the balance of care model proved unsuccessful, and enjoyed very little practical use.

2.2   Performance indicators

The NHS performance indicator (PI) initiative was launched in 1983 with the publication of a series of indigestible reports comparing the financial performance of Regional and District Health Authorities. The stated purpose of the initiative was "to help [managers] assess the efficiency of the services for which they are responsible" (DHSS, 1983). It was immediately clear that the vast quantity of data contained in the reports would have little impact if it was left in written form, so the next edition of performance indicators was distributed in machine readable format, suitable for use with LOTUS software (DHSS, 1985). Over 400 indicators were included, as summarized in Table 1. The emphasis was on measures of inputs (in particular personnel) and process (such as length of stay), and there were very few measures of outcome. The system has since been augmented by many additional indicators.

7

1. Epidemiological variables (3 in total)
Low birthweight rates
Infectious diseases notification rate

2. Resource provision (100 in total)
Staff (by category) per inpatient case
Nurses per bed (by specialty)
Nurses in relation to relevant catchment population (non-acute services)
Paramedical staff (by category) in relation to catchment population
Medical manpower (by category) in relation to catchment population
Support staff (by category) in relation to workload
Provision of beds in relation to catchment population (non-acute services)
Theatre sessions per bed (by surgical specialty)
Ambulance staff (by category) in relation to workload
Provision of land in relation to resident population
Provision of buildings in relation to resident population
Works staff in relation to building area

3. Resource quality (52 in total)
Percentage of non-medical staff in each staff category
Percentage of nursing staff by level of training (by specialty)
Ratio of qualified to unqualified paramedical staff (by category)
Disposable land as percentage of total
Building area in relation to bed provision

4. Resource costs (93 in total)
Costs in relation to resident population
Costs in relation to inpatient cases
Staff costs as a percentage of revenue expenditure
Staff costs by service per unit of workload (e.g. inpatient case) by category
Staff costs in relation to resident population (community services)
Cost per whole time equivalent member of staff (by service and category)
Total costs by service per unit of workload
Building maintenance costs in relation to area

5. Process variables (165 in total)
Admission rates (by service, specialty and certain conditions)
Average length of stay (by specialty and certain conditions)
Distributions of length of stay (non-acute services)
Throughput (by specialty)
Expected throughput (by specialty)
Turnover interval (by specialty)
Waiting list in relation to catchment population (by surgical specialty)
Days needed to clear waiting list (by surgical specialty)
Percentage of cases not operated on (by surgical specialty)
Contact rates with certain client groups
Special care baby unit admissions in relation to total births

**Table 1: Categorization of National Health Service performance indicators
(Source: Smith, 1990)**

The PI initiative appeared at first to have very little impact on the service managers at whom it was ostensibly aimed (Jenkins *et al*, 1988), although the Government had made strenuous efforts to market the package (DHSS, 1988). In fact it soon became clear that, far from offering service managers the opportunity to secure control of their resources, the principal objective of the initiative was to service the NHS system of performance review, and to enable the centre to secure control of the periphery.

The government promoted this use of the PIs vigorously. The aim was "to build on good performance, challenge and analyze poor performance and generally to provide a clearer definition of the objectives of expenditure and the expected results" (Trumpington, 1987). The computerized forms of the performance indicators were distributed with a computer program which enabled users to select items for investigation, and to examine with some ease the performance of their organization in relation to that of others. Thus, some attempt was made to make the package accessible to service managers. However, to help the review process it was necessary to be able to scan the PIs selectively in order to focus on particular areas of concern. Scrutiny of performance on all PIs would clearly be infeasible. The need was for a methodology that "identified extreme performance, identified possible causative factors and was consistent between authorities" (Bowen and Payling, 1987). As a result, the Operational Research Service of the DHSS developed an expert system using the CRYSTAL shell to assist the process of performance review (DHSS, 1987). This system sought to automate the extremely complex process of sifting and exploring the vast range of indicators by focusing on aberrant behaviour and seeking to explain why it arose.

The process underlying the expert system is very simple. A set of "first line" performance indicators is defined, usually reflecting intermediate outputs of the NHS, such as waiting times. These PIs are then examined for extreme performance, most especially, one might imagine, what is judged to be extremely poor performance. Where unsatisfactory performance amongst first line indicators is found, a more detailed investigation of "second line" indicators is instigated. This might entail the investigation of resource provision and the use made of those resources, such as bed throughput and length of stay. The outcome should be an explanation of aberrant behaviour, with an implicit prescription for a strategy to improve matters.

9

Clearly many judgements are required to implement the expert system. The PIs to be investigated must be selected, and definitions of high and low performance developed. Models must then be developed of the links between the first and second line indicators. In practice, as Bowen and Payling (1987) show, the models underlying the expert system are very rudimentary, representing no more than the applied common sense that Boldy and Clayden found dominated health care OR. For example, high waiting times might be caused by low provision of beds or theatres or doctors. If resource provision is not unusual, then there might be low throughput, which can be investigated by examining PIs related to turnover interval and lengths of stay. This process is no more than that which any diligent manager would apply to a problem identified with local services. The need for the expert system arises because of the central government control problem. It feels it needs to scrutinize a vast amount of data pertaining to a large number of organizations, and therefore requires a mechanism for isolating the most important apparent problems. In addition, it must be seen to be acting fairly and consistently between jurisdictions.

The expert system is perhaps the most ambitious application of modelling in the NHS, seeking to secure control of the entire organization. The absence of any theory underlying the structure of the model and its intellectual parsimony should not disguise its large scale ambitions. From the perspective of this paper, its importance lies not so much in its usefulness, as in the insight it offers into the philosophy of central control it makes explicit. The model lays bare the manner in which the central government envisages PIs being used in the review process. There is informal evidence that at least some of the 14 NHS Regions are actively using the expert system. However, there are only a handful of enthusiastic users amongst the Districts.

Performance indicators have now become an integral part of the system of performance review within the NHS. There is increasing evidence that the NHS Management Executive is being held to account by Parliament on the basis of PIs (Committee of Public Accounts, 1990). The culture of PIs is therefore permeating the NHS, as the Management Executive in turn holds its devolved Regions to account with the same indicators. There is also evidence that - at the lower end of the hierarchy - the providers of NHS services are being held to account by PIs in the terms of their contracts with purchasers (Smith,

10

forthcoming).

2.3   The internal market

In 1990 the UK Government introduced major reforms to the NHS. The most important innovations related to the provision of hospital and community services. In the past these had been delivered directly by District Health Authorities. Under the reforms, DHAs purchase their services from a variety of sources. Some services are still provided directly by the DHAs themselves. However, a rapidly increasing proportion is provided by NHS Trusts, which are run independently of the DHA. In addition, some large family practitioner practices purchase services in a similar manner for their patients, reducing the scope of the DHA's responsibilities. In these circumstances, therefore, the mechanism of control over service managers has changed from direct line management to explicit contract between purchaser and provider. The new arrangement seeks to introduce an "internal market" into the NHS, whereby the functions of purchase and provision of services are divorced. This is intended to "encourage local initiative and greater competition [thereby] improving the choice and quality of the services offered and the efficiency with which those services are delivered" (Department of Health, 1989, p22).

The inspiration for the reforms is clearly the model of market competition adumbrated by economists (Culyer and Posnett, 1990). The holy grail of economists is the notion of Pareto efficiency. Within a theoretical framework it is possible to show that, under a set of restrictive assumptions, a perfectly competitive market is one means of securing an efficient outcome (Ng, 1983, Chapter 2). In practice, of course, many imperfections conspire to frustrate the free operation of markets. In particular, in the health sector there is no obvious objective function, such as profit maximization, for "producers" in the market; "consumers" have very imperfect information about the efficacy of the available products (or treatments); there are large economies of scale in the provision of hospital services, leading to natural monopolies; and notions of equity as well as efficiency appear to be important considerations in forming society's preferences.

Nevertheless, the UK Government had sufficient faith in the economic model of

11

competition to embark, without experimentation, on a massive transformation of an organization that, in international terms, already appeared to be comparatively efficient (Appleby, 1992). Here, then, is an example of a theoretical model fundamentally influencing government policy in a way which most management scientists would find inconceivable. At the time of writing it is too early to make definitive judgements on the effects of the reforms. However, Mullen (1992) is typical of many commentators in suggesting that few of the government's objectives, as set out in the 1989 White Paper, are being met. Notwithstanding these preliminary findings, the government appears committed to an acceleration of the process of creating independent Trusts.

3.   DISCUSSION

The three developments described here are of interest because they signify the ascendancy of different approaches to public sector management. They are now discussed in turn.

3.1   The balance of care model: planning the system

The balance of care model reflected the traditional, rationalistic operational research approach to planning health care. In presuming that objectives could be expressed in terms such as cost minimization or benefit maximization, that value judgements could be quantified, and that "experts" could define ideal levels of care, the model was the apotheosis of rational health care planning. Its failure to have anything other than a marginal impact is therefore instructive.

Flood and Jackson (1991) categorize problem contexts according to the complexity of the system under scrutiny, and the political structure within which the participants operate. Within this "system of systems" framework, health service planning is clearly a complex problem with pluralistic relationships between participants. Flood and Jackson conclude that in such systems it is likely to be methodologies such as interactive planning and soft systems that are most appropriate. The balance of care approach represents the antithesis of such methodologies, and is more suitable for simple problems with unitary relationships. Thus,

if one accepts the Flood and Jackson view, it seems that the architects of the model were in some senses seeking to wish away both the complexity of the health care system and the richness of the political pressures at work.

In a survey of OR applied to health services roughly contemporary with the balance of care project, Rosenhead (1978) pointed to the dangers of applying a traditional OR approach to health services planning. His conclusion was that "we should prize approaches which: make reduced demands on data; reject optimisation in favour of coordination; accept uncertainty and try to keep options open; are not restricted to hierarchical deduction, but facilitate participation; [and] do not attempt a technocratic abolition of politics". The balance of care fails on all five of these counts.

Its appetite for data is voracious. Even apparently straightforward items, such as the size of client groups, are unknowable. Many persons will be members of more than one group, and the size of a group will to some extent be determined by the nature of the services provided. Thus, if the rationalistic approach of the model is pursued to its logical conclusion, there should be feedback from the output of the model (resource allocations) to the inputs (client groups). Similarly, it is meaningless to ask anyone to specify ideal levels of care without reference to the resources available. Like client group size, levels of care should also be considered endogenous to the model. Finally, of course, the far more nebulous concepts relating to worth have to be quantified in a systematic and non-contradictory manner. Even in laboratory conditions, with very simple problems, experimental economists find that subjects are unable to remain consistent in their judgements (Hey, 1991). The balance of care philosophy modeller is therefore highly unlikely to obtain anything other than an arbitrary and inconsistent selection of value judgements.

Although the authors of the model advocate its use to simulate a variety of options, some sort of objective function is intrinsic to its functioning. Yet the objectives underlying the health services are far from clear. Certainly concepts such as minimizing costs and maximizing effectiveness are important. However, most popular debate on health services also reflects a strong concern with considerations of equity, the principle on which the NHS was created. Unfortunately, neither economists nor other commentators have yet been able

13

to develop convincing concepts of equity equivalent to the notions of Pareto efficiency that have found such widespread favour. Nevertheless, it seems important that any model of resource allocation should pay some attention to equity considerations. In seeking to pursue some notion of allocative efficiency, the balance of care model does pay some attention to fairness. However, the balance attained between client groups is only a by-product of the optimization model, and is not incorporated into the objective function.

Underlying the balance of care model is the idea that, if the data problems could be overcome, the model would be a good reflection of reality. Uncertainty can only be accommodated by tedious and arbitrary sensitivity analysis, and no measures of the robustness of solutions is produced. Yet uncertainty is intrinsic to the health services. Technologies change with remorseless frequency. As technologies change, so do citizens' expectations and demands. To the health service planner, uncertainty is the key concept to be managed and accommodated. The balance of care model gives little attention to such dynamic preoccupations.

By the admission of its architects, the complexity of the model precludes participation by all but a handful of cognoscenti. A typical manager will probably have neither the time nor the training to comprehend the assumptions underlying the model. Comprehension amongst consumers or their representatives is even less likely. As a result, the model is intrinsically "top down". If its implications are implemented, the values and prejudices of the modellers, however well intentioned, will prevail. Two possible outcomes will emerge. The approach adopted by the advocates of the model is that modellers should seek to infer values from existing behaviour, with all its anomalies and contradictions. A less benign scenario is that the centre uses the model to promulgate its own agenda, regardless of popular preferences. Thus, because of the non-participative culture implicit in the model, its methodology must be intrinsically conservative or authoritarian. Evidence from the management control and Soviet literature suggests that both cultures are severely dysfunctional (Emmanuel, Otley and Merchant, 1990).

More generally, the model ignores the complex political context within which the health and personal social services operate. For example, the model makes no

14

acknowledgement of the political reality that health services are delivered by the NHS, a centralized organization, and personal social services by local government, a devolved level of government with local accountability. In many areas, administrative boundaries for these two organizations are not even coterminous. Similarly, the model simply ignores the perpetual political tensions that exist within the NHS, brought about by the multiplicity of interested parties, such as managers, doctors, nurses, patients, voluntary groups, carers and the central government, each with their own distinct priorities.

Thus, even if a central or local administration was favourably disposed towards the model, it is unlikely that it would produce helpful results. More importantly, however, the philosophy underlying the model is likely to be profoundly unattractive to politicians. If the model is to secure change, someone must be made accountable for implementing its findings. Let us suppose, for example, that the central government uses the model at a national level. The Secretary of State might then be the accountable person. Now, even if that person feels the model is a perfect reflection of society's preferences, he or she is unlikely to want to be associated with it for two reasons. First, the specification of the model requires explicit judgements to be made about the importance of fulfilling the demands of different segments of the community. And second, the output from the model includes very detailed allocations of resources between competing claims. The model therefore lays bare prejudices and decisions in a way which no politician could tolerate: in Benjamin Disraeli's words, "finality is not the language of politics". The astute politician wishes to delegate blame and keep options open as much as possible, and the centralized culture of the balance of care model is completely antipathetic to such objectives.

This paper is principally concerned with the model developed in the 1970s, and its failure to have any significant impact on operational management. However, it is interesting that the Department of Health (1991b) subsequently developed a second version of the Balance of Care model. The changes from the original model are noteworthy as they serve to highlight some of the weaknesses of the original model, most especially the need for expert intermediaries to specify parameters and run the model. In particular, the second model, for use on personal computers: addresses only long term care for the elderly; is menu driven, and designed for use by operational managers; and is provided with many of the parameters

already estimated - only local demographic data and service provision are required in the first instance.

3.2   Performance indicators: securing control

The performance indicator initiative reflects a different culture: that of management accounting. The advocates of PIs appear to have in mind the application of cybernetic principles to complex human organizations (Hofstede, 1981). PI data are used to scan the activity of devolved units. As a result, targets are set, either implicitly (if typical behaviour is used as the benchmark) or explicitly. Variances from such targets are noted. And strategies for remedial action can be developed by examining more detailed data. Organizational behaviour is thereby affected, and the outcome can be monitored with PIs. So the feedback process continues, reflecting very closely the mechanistic management control philosophy advocated by Anthony and Young (1984).

Few would question the desirability of enhancing the scope of the information available to NHS management and their political controllers. However, great care must be taken in the use to which such data are put. The accounting and Soviet literature which documents the potentially distorting affects of cybernetic control systems is extensive (Briers and Hirst, 1990; Nove, 1980), and Smith (forthcoming) notes at least seven ways in which excessive attention to PIs might influence public sector managerial behaviour in unintended ways. In particular, it might encourage the following:

(a)   tunnel vision: concentration on areas included in the PI scheme, to the exclusion of other important areas;

(b)   suboptimization: the pursuit by managers of their own narrow objectives, at the expense of strategic coordination;

(c)   myopia: concentration on short term issues, to the exclusion of long term considerations, which may only show up in PIs in many years' time;

16

(d)   convergence: an emphasis on not being exposed as an outlier on any PI, rather than a desire to be outstanding;

(e)   ossification: a disinclination to experiment with new and innovative methods;

(f)   gaming: altering behaviour so as to obtain strategic advantage, for example by deliberately performing poorly in one year so as to secure easily achievable targets in future;

(g)   misrepresentation: including "creative" accounting and fraud.

There is evidence that, as the performance indicator culture begins to bite, some of these phenomena are becoming evident. Thus The Guardian (1992) reports that, in accordance with central government targets, as set out in the Patient's Charter, DHAs are rapidly reducing the numbers of patients waiting more than two years for elective surgery. However, this appears to be at the expense of an increase in the average wait for all surgical patients, particularly those awaiting more serious surgery. This side effect of management by an imperfect and incomplete PI set may or may not be desirable: however, it is almost certainly unintended.

Similarly, implicit in the PI culture - as expressed in the expert system - is the notion that there is some model level of service, representing best practice. In fact, it is highly likely that the emphasis on poor performers is likely to encourage managers to adopt typical practice, rather than best practice. There is clearly a strong disincentive to be exposed as an extreme performer. This being the case, the PI initiative has important implications for resource allocation at the local level. A locality will be encouraged to adopt regional or national norms for waiting times, resource provision and throughput, even if this contradicts local preferences.

Nevertheless, in spite of its manifest dangers, the PI culture has permeated the NHS in a way that the apparently more comprehensive and systematic planning culture never did. In seeking an explanation, one must first note the considerable amount of political support

17

for PIs. This is illustrated by the high profile given to the initiative by government ministers (Trumpington, 1987), and, most dramatically, by the important role they play in John Major's flagship, The Citizen's Charter (Department of Health, 1991a). The reason why PIs are attractive to politicians is clear. Notwithstanding its potential pitfalls, the initiative appears to enhance the control that the centre has of devolved units. But it also enables the centre to devolve blame for poor performance down to the unit level, by allowing the centre to point to other areas which appear to produce better results. The ability to focus on particular aspects of good or bad performance, rather than search for some systemic optimum is clearly attractive, as it suits the nature of political discourse. PIs can be invoked when they support rhetoric and ignored otherwise. From the politician's viewpoint, there is no need to have a comprehensive, balanced coverage of activity, so long as there is enough diversity to control devolved management, at least along some dimensions of performance. In addition, PI data offer explanations or excuses for poor performance, thus facilitating the delegation of blame, which can be shifted to local management or uncontrollable environmental factors (Armstrong, 1991).

This suggests that the politician's principal concern is not with optimizing performance (however defined) but with securing control. As Thain and Wright (1992) note, this preoccupation has become evident in the central government's public expenditure planning process, in which the control of expenditures is given much higher priority than the pursuit of any sort of allocative efficiency related to public expenditure. Similarly, the long history of the central government's relentless and arbitrary campaign against local government (Travers, 1986) is probably a symptom of the frustration that the centre feels at the lack of control it exercises over local government, rather than a desire to secure a more socially desirable pattern of service provision by local government.

The associated expert system is not so much a model of the health services, as a model of the government's control mechanism, as it makes explicit one interpretation of the control mechanism. Its philosophy conforms to the policy of "mixed scanning", long advocated by the planning theorist Etzioni (1968, Chapter 12) in preference to the rational-comprehensive approach for messy, poorly structured planning problems. Certainly the expert system's systematic codification of the review process might be helpful for low

level, routine scrutiny of devolved units. However, because it forces dynamic consistency on users, and obliges them either to make explicit their own control procedures, or to accept some expert's interpretation, it is unlikely to prove attractive to politicians. The model is therefore unlikely to be given the same sort of political support given to the PI initiative as a whole.

3.3   The internal market: avoiding blame

The balance of care model sought to model in great detail the existing system of health and personal social services, relying on expert specification and existing practice for calibration. The economic model of markets, on which the 1990 reforms are based, is predicated on the belief that such calibration is futile, and that only the micro-decisions of many competing purchasers and suppliers can deliver an efficient (or socially optimal) solution. Implicit in the market model is the belief that the information requirements of the planning model are impossibly large; and that - because of this - various players have the incentive and opportunity to distort information to their advantage, with severely dysfunctional consequences (Rose and Miller, 1992). Thus, for example, it might be argued that the "experts" used to calibrate the balance of care model may, knowingly or otherwise, distort their assessments of ideal levels of care in order to further their own objectives.

The traditional market model is based on a large number of assumptions, of which some of the most important are:

(a)   the existence of a large number of buyers and sellers, such that no one individual can affect either prices or volume;

(b)   free entry to and exit from the market;

(c)   information freely available to all participants;

(d)   zero transaction costs.

19

In practice, none of these conditions holds in the health services. Most purchasing DHAs are virtually monopoly buyers in their locality, and many hospital services are natural monopoly providers, with the ability to inhibit entry of competitors to the market. Unlike supermarkets, NHS organizations cannot choose to decline to provide a service. They are obliged to attend to the needs of all patients, however inconvenient, and must select what they consider to be the most cost-effective mode of care amongst the limited choices available. The availability of information is limited in many respects, and susceptible to distortion by both providers and purchasers. And the transaction costs to purchasers of writing and monitoring contracts are potentially very large.

However, the fact that none of the conditions for securing an efficient competitive market equilibrium is even remotely satisfied did not deter the UK Government from wholesale reform in 1990. It therefore seems likely that it is not the predicted outcome of the market model that is attractive, but some of its intrinsic characteristics. The viewpoint of this paper is that the concept of accountability holds the key to understanding the model's success in securing political favour. Indeed in some ways the internal market reforms can be viewed simply as an extension of the PI culture. By exposing comparative performance, PIs introduce pseudo-competition between units. The internal market seeks to introduce more tangible competition, with providers of services being held to account by legal control mechanisms (contracts) rather than internal PI systems.

Again, it is clear why this development is attractive to politicians at the centre. True accountability has two components: the rendering of an account by the agent; and the holding to account by a principal. In his evidence to the Public Accounts Committee (1990), the Chief Executive of the NHS demonstrates that, while PIs can deliver the first component, he concedes that the only sanction he has available for poor performers is to give them "a hard time". Such nebulous sanctions are clearly inadequate for a government committed to securing strong control of the NHS. The internal market completes the accountability structure by offering the centre the opportunity to dispose of incompetent, inconvenient or unlucky management. Moreover, by appearing to delegate decisions to a very local level, the internal market allows the centre to absolve itself of responsibility for detailed allocation of resources. The delegation of blame - facilitated by the PI initiative - is promoted still further

20

by the internal market. Thus the reforms seek to maximize the centre's ability to hold to account, and minimize the danger of it being held to account itself.

Of course, some proponents of the virtues of competition would claim that the outcome may also be the enhancement of managerial efficiency. In particular, they might argue that the fear of losing contracts or becoming insolvent will drive management in providing organizations to search vigorously for schemes to enhance outcome and reduce costs. Equally, a sceptic might argue that, given the market imperfections noted above, managers will instead divert their attentions to retaining their monopoly position, or manipulating information to their advantage. Whichever turns out to be true, the fact that the economic model predicts an enhancement of managerial efficiency probably increased the attractiveness of the reforms to the government, but is probably secondary to the principal attraction of the reforms: the enhanced control they offered to the centre.

4. CONCLUSIONS

Within the context of health services, this paper has sought to explain why the traditional operational research modelling paradigm has failed to have an enduring impact on service delivery, while the disciplines of accounting and economics have contributed to profound changes in the culture of the NHS. All three of the developments here take a standard model "off the shelf" and apply it the NHS. However, the ways in which the models are implemented differ fundamentally. The balance of care model uses the archetypal OR technique, a mathematical programme, and seeks to fit it to the NHS system. The PI initiative uses accounting methods which are so general and free of theory that they need little adaptation to be used immediately. The internal market reforms take the standard economic model of markets, and adjust the health system to conform to the model. A summary of the characteristics of each of the three initiatives is given in Table 2.

21

MODEL	DISCIPLINE	MANAGERIAL STYLE	PHILOSOPHY OF CONTROL
Balance of care	Operational research	Planning	Open loop PRESCRIPTION
Performance indicators	Accounting	Bureaucratic	Closed loop MONITORING
Internal market	Economics	Market	Closed loop CONTROL

Table 2: Characteristics of the three models

The OR modelling approach, as embodied in the balance of care model, reflects the rational-comprehensive planning philosophy. It seeks to model in great detail some sort of Platonic notion of the health care system in which demand and resources appear from nowhere, and simply need to be matched to each other according to some divine concept of optimality. Inherent in the approach is an "open loop" notion of control, in which the system is optimized at the start of the planning period, and no explicit attention is paid to the inevitable need to reconsider strategy as time unfolds. Numerous technical concerns about the feasibility of modelling such a complex and messy process as NHS resource allocation have been raised. Moreover, and more fundamentally, the approach chooses to gloss over entirely the political context within which the decisions emerging from the model are to be implemented, in particular the reality that some manager or politician will be held accountable for those decisions. We suggest that, far from being a strength, the stark clarity of the model makes it unacceptable to the accountable person. Given the client-oriented emphasis of OR (Royston, 1992) this finding is perhaps surprising. However, instead of offering control of the system, the balance of care model exposes the centre to unwanted scrutiny, and offers few degrees of freedom in searching for excuses or delegating blame. In short, the model is over-specified, and its intricate interlocking structure is its key weakness.

The PI initiative emphasizes the bureaucratic approach to management. It is based on a "closed loop" model of control, with an emphasis on continual monitoring of the system so that aberrations can be detected and appropriate remedial action taken. By offering the illusion of control, however unfounded, it appeals to the politician's priorities. Moreover, in contrast to the straitjacket imposed by the balance of care model, it offers numerous degrees of freedom to the user. The lack of structure of the PI model is therefore its main strength. For example, it does not specify modes of care, so such matters can be left to local discretion, unless the centre feels it is useful to issue guidelines, in which case it can monitor care processes with ease. It therefore allows the centre to delegate blame to devolved units when convenient, to blame uncontrollable environmental factors when not, and even to exercise direct control if necessary. In short, it promotes the accountability of devolved units to the centre. The PI model also benefits from the property of "selective ignorability": the centre can choose to ignore those aspects of the model that do not suit its purpose, and can concentrate on aspects which do.

The internal market reforms are in some ways a paradox. The "invisible hand" of markets is advanced by economists as a solution to the insurmountable problem of collecting and processing the information required to arrive at an efficient allocation of resources. Yet, in order to write and monitor contracts, purchasers require huge volumes of hitherto unavailable information. The market model therefore seems completely inappropriate to the health sector. In spite of this, the NHS reforms are being implemented with great haste. This paper has argued that the market model is attractive for politicians not so much because of its efficiency implications, but because - by introducing potential sanctions against local management - it enhances still further the control of the centre, and - by devolving decision making - it facilitates the delegation of blame. It therefore completes the closed loop control model by (perhaps illusorily) enhancing the ability to take effective remedial action.

The lesson for OR from this story is simple. It is that large scale problems attract great popular, and therefore political, interest. As a result, the modeller of such systems must pay great attention to the pattern of accountability implicit in the system being modelled. If the model is accepted, the accountable person will want (a) to ensure that its findings are implemented at all levels and (b) to deflect blame for any problems that arise from

23

implementation. The consummate politician would wish to maximize control while minimizing responsibility. If large scale OR models are to be successful in influencing policy they must recognize this phenomenon.

The OR community has recognized the need for more politically sensitive modelling for some time (Rosenhead, 1989). However, many of the existing techniques for modelling complex, pluralistic systems depend on convergence toward consensus and a reconciliation of conflicting viewpoints. The argument put forward here is that even this philosophy may be flawed. If OR modellers are to contribute usefully to large scale problems, then their models must acknowledge the priorities of those who will be held responsible for the solutions they propose. At present, political realities, in which accountability plays such a key part, are better served by the models of accountants and economists.

Notwithstanding the limitations documented in this paper, all three of the approaches described here contribute valuable insights into the management of complex systems. The OR approach probably offers the closest technical representation of the problem being addressed. The PI initiative reflects the need to keep the system under continual surveillance. And the market reforms emphasize the need to be able to take effective action. Almost certainly, dogmatic adherence to only one of the approaches is likely to lead to disappointing results, and the most successful approach to managing the NHS is likely to involve a judicious mix of the principles of all three types of approach. Stafford Beer (1966, p495) rather hopefully commented that "this combined operational research and cybernetic technique has been evolved to deal with the kind of large scale decision described here. It ought to be appropriate to most governmental decision-taking." Twenty-five years later, the NHS is only rather hesitantly moving in the direction he espoused.

The policy implication of this paper's conclusions is therefore that OR practitioners seeking to become involved in politically sensitive large scale problems should be prepared to adopt some of the principles of related disciplines, and to consider a hybrid approach to aiding control. If instead they adopt the traditional OR model, then their intervention is likely to be unsuccessful. Practitioners should moreover think carefully about the implications of their involvement in the policy domain. If their intervention is successful, in the sense of

securing political approval, it may be that they are simply strengthening the role of the centre, without contributing to a better social outcome. Perhaps the principal role of traditional large scale OR models in this respect is therefore to make policy makers aware of the technical structure of the problem being confronted, and to alert them to the possibility of securing improvements in the system, as well as merely controlling it.

References

Anthony, R. N. & Young, D. W. *Management control in nonprofit organizations,* third edition, Homewood: Richard D. Irwin, 1984.

Appleby, J. *Financing health care in the 1990s*, Buckingham: Open University Press, 1992.

Armstrong, P. Contradictions and social dynamics in the capitalist agency relationship, *Accounting, Organizations and Society*, 1991, 16, 1-25.

Beer, S. *Decision and control*, Chichester: Wiley, 1966.

Boldy, D. and Clayden, D. Operational research projects in Health and Welfare Services in the United Kingdom and Ireland, *Journal of the Operational Research Society*, 1979, 30, 505-511.

Boldy, D., Canvin, R., Russell, J. and Royston, G. Planning the balance of care, in D. Boldy (ed) *Operational research applied to health services*, London: Croom Helm, 1981.

Bowen, B. and Payling, L. Expert systems for performance review, *Journal of the Operational Research Society*, 1987, 38, 929-934.

Briers, M. and Hirst, M. The role of budgetary information in performance evaluation, *Accounting, Organizations and Society*, 1990, 15(4), 373-398.

Committee of Public Accounts *Maternity services*, London: HMSO, 1990.

Coverdale, I. and Negrine, S. M. The balance of care project: modelling the allocation of health and personal social services, *Journal of the Operational Research Society*, 1978, 29, 1043-1054.

Culyer, A. J. and Posnett, J. Hospital behaviour and competition, in A. J. Culyer, A. K. Maynard and J. W. Posnett (eds) *Competition in health care: reforming the NHS*, Basingstoke: Macmillan, 1990.

Davies, H. and Davies, R. A simulation model for planning services for renal patients in Europe, *Journal of the Operational Research Society*, 1987, 38, 693-700.

Department of Health *The Patient's Charter*, London: Department of Health, 1991a.

Department of Health *The Balance of Care System: introduction to release 2.0*, London: Department of Health, 1991b.

Department of Health *Working for patients*, CM555, London: HMSO, 1989.

Department of Health and Social Security *Performance indicators: National summary for 1981*, London: DHSS, 1983.

Department of Health and Social Security *Performance indicators for the National Health Service, guidance for users*, London: DHSS, 1985.

Department of Health and Social Security *The performance indicator analyst, version 1.0*, London: DHSS, 1987.

Department of Health and Social Security *Comparing health authorities*, London: DHSS, 1988.

Emmanuel, C., Otley, D. and Merchant, K. *Accounting for management control, Second edition*, London: Chapman and Hall, 1990.

Etzioni, A. *The active society: a theory of societal and political processes*, New York: The Free Press, 1968.

Flood, R. L. and Jackson, M. C. *Creative problem solving. Total systems intervention*, Chichester: John Wiley & Sons, 1991.

The Guardian, Big rise in queues for hospital treatment, *The Guardian*, 13 May 1992.

Gibbs, R. J. The use of a strategic planning model for Health and Personal Social Services, *Journal of the Operational Research Society*, 1978, 29, 875-883.

Gibbs, R. J. Healthcare resource allocation models, in D. Boldy (ed) *Operational research applied to health services*, London: Croom Helm, 1981.

Hey, J. D. *Experiments in economics*, Oxford: Blackwell, 1991.

H.M. Treasury *Public expenditure analyses to 1994-95: statistical supplement to the 1991 Autumn Statement*, Cm1920, London: HMSO, 1992.

Hofstede, G. Management control of public and not-for-profit activities, *Accounting, Organizations and Society*, 1981, 6(3), 193-211.

Jenkins, L., Bardsley, M., Coles, J. and Wickings, I. *How did we do? The use of performance indicators in the National Health Service*, London: CASPE Research, 1988.

McDonald, A. G., Cuddeford, G. C. and Beale, E. M. L. Mathematical models of the balance of care, *British Medical Bulletin*, 1974, 30, 262-270.

Mullen, P. *The impact of health service systems on health care*, Paper presented to Operational Research Society Annual Conference, Birmingham, September 1992.

Ng, Y. *Welfare economics. Introduction and development of basic concepts*, London:

Macmillan, 1983.

Nove, A. *The Soviet economic system*, London: Allen and Unwin, 1980.

Resource Allocation Working Party *Sharing resources for health in England*, London: HMSO, 1976.

Rose, N. and Miller, P. Political power beyond the state: problematics of government, *British Journal of Sociology*, 1992, 43, 173-205.

Rosenhead, J. Operational research in health services planning, *European Journal of Operational Research*, 1978, 2, 75-85.

Rosenhead, J. (Editor) *Rational analysis for a problematic world*, Chichester: Wiley, 1989.

Royston, G. H. D. A systematic framework with illustrative UK examples for applied health OR/MS, in M. K. Chytil, G. Duru, W. V. Eimeren and D. Flagle (eds) *Fifth International Conference on System Science in Health Care*, Prague: Omnipress Publishing, 1992.

Smith, P. The use of performance indicators in the public sector, *Journal of the Royal Statistical Society, Series A*, 1990, 153, 53-72.

Smith, P. Outcome-related performance indicators and organizational control in the public sector, *British Journal of Management*, forthcoming.

Thain, C. and Wright, M. Planning and controlling public expenditure in the UK, Part II: the effects and effectiveness of the survey, *Public Administration*, 1992, 70(1), 193-224.

Travers, T. *The politics of local government finance*, London: Allen and Unwin, 1986.

Trumpington, Baroness OR in government, *Journal of the Operational Research Society*, 1987, 38, 907-912.