

CONTROLLING NHS EXPENDITURE: THE IMPACT OF LABOUR'S NHS WHITE PAPERS

Jon Sussex

Office of Health Economics



CONTROLLING NHS EXPENDITURE: THE IMPACT OF LABOUR'S NHS WHITE PAPERS

Jon Sussex
Office of Health Economics

December 1998



Office of Health Economics
12 Whitehall
London SW1A 2DY

© December 1998. Office of Health Economics. Price £5.00

ISBN 1 899040 75 7

Printed by BSC Print Ltd, London.

About the Author

Jon Sussex is the Associate Director of the Office of Health Economics (OHE). His current areas of research concern the organisation of, and use of incentives in, the UK National Health Service. Prior to joining the OHE, Jon worked for several years as a management consultant in the NHS. Before that, he was HM Treasury's economic adviser on health and personal social services.

Office of Health Economics

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:

Chairman

Professor Lord Maurice Peston – Queen Mary and Westfield College, University of London

Members

Professor Michael Arnold – *University of Tübingen*

Dr Michael Asbury – *Zeneca Pharmaceuticals*

Mr Michael Bailey – *Glaxo Wellcome and President of the Association of the British Pharmaceutical Industry*

Mr Richard Bailey – *Eli Lilly and Company*

Professor Tony Culyer – *University of York*

Professor Patricia Danzon – *The Wharton School of the University of Pennsylvania*

Dr Trevor Jones – *Director General of the Association of the British Pharmaceutical Industry*

Dr Nancy Mattison – *Director, Pharmaceutical Partners for Better Healthcare*

Professor Sir Michael Peckham – *University College, University of London*

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 Christopher Bulpitt – *Royal Postgraduate Medical School, Hammersmith Hospital*

Professor Martin Buxton – *Health Economics Research Group, Brunel University*

Professor Stewart Cameron – *Emeritus Professor of Renal Medicine, United Medical and Dental Schools, London*

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

1. Introduction	7
2. Cash Limits in the NHS 1998/99	10
2.1 Public Expenditure Control since 1948	10
2.2 Controlling NHS Prescribing Expenditure	13
3. The 1997/98 White Papers – Extending the NHS Cash Limit	19
4. The 1997/98 White Papers – Abolishing GP Fundholding	24
5. Incentives for GPs	27
6. The Result: Calling the Budget Bluff?	33
7. So What?	39
Appendix	42
References	44

1 INTRODUCTION

The English, Scottish and Welsh National Health Service (NHS) White Papers published by the government in December 1997 and January 1998, have changed the tone of NHS policy. Co-operation is to replace competition; there is to be a statutory requirement to provide good quality health care; and performance benchmarking is expected to succeed where market forces failed in producing efficiency gains. The consultation paper 'Fit for the future' published by the Department of Health and Social Services in May 1998 implies that the same changes will apply in Northern Ireland too. How far the practical operation of the NHS is altered by the White Papers' proposals remains to be seen, but two changes that clearly will occur, and which are the focus of this paper, are:

1. the abolition of fundholding by individual practices of general medical practitioners (GPs) and their replacement by various forms of collective GP involvement in commissioning;

2. the inclusion for the first time within the cash limit applied to most NHS spending, of all of the cost of medicines prescribed by GPs. Hitherto, although the prescribing expenditure of fundholding GPs had been brought within the overall NHS cash limit, the prescribing of non-fundholders had remained non-cash limited.

The aim of the first of these measures appears mainly to be to avoid a 'two-tier NHS', where the patients of fundholding GPs may have been able to obtain referrals for specialist diagnosis and treatment more rapidly than could the patients of non-fundholders. A secondary aim is to contribute to cutting bureaucracy and hence costs, by reducing the number of parties negotiating with health care service providers and being invoiced by them. GP fundholding has spread widely, so that by 1997/98 over half of the UK population was registered with fundholding GPs. Abolition of this scheme from 1 April 1999 (but a year later in Northern Ireland) is likely to have effects beyond those apparently sought by the government, however.

In particular, it will weaken the incentives on ex-fundholding GPs to restrain their prescribing costs and those non-prescribing costs that were included in their fundholder budgets. At the same time, as will be argued in section 5 of this paper, it is currently unclear whether there will be a significant offsetting strengthening of the incentives on non-fundholding GPs (i.e. all GPs once the new policies are in opera-

tion) to restrain their total expenditures. Thus a pre- and post-White Papers comparison implies that the incentives on GPs to control expenditure will weaken in total. It remains to be seen whether that weakening will be significant.

The extension of the NHS cash limit to encompass the total Family Health Services (FHS) medicines bill, i.e. including for the first time the cost of medicines prescribed for NHS patients by non-fundholding GPs, may be intended to serve two main purposes. Firstly, it may be expected by the Treasury to tighten its control over NHS costs by bringing a further six per cent of total NHS expenditure under the (supposedly) fixed annual cash limit. (In 1997/98, in England, non-fundholding GPs' prescribing cost £1.9 billion and fundholding GPs' £2.2 billion. These sums equate respectively to just under and just over six per cent of total NHS expenditure.) Secondly, extending the cash limit takes away one possible distortion of rational clinical choices, by removing the incentive for non-fundholding GPs to prescribe medicines for their patients (for which expenditure is not cash limited) rather than refer them to specialist hospital or community health care services (for which expenditure is cash limited). It is unclear, though, whether this potential distortion to treatment choices has been significant in practice.

In this paper, I shall first describe the way in which NHS spending has been controlled until, and including, the 1998/99 financial year. In that context I shall then go on to assess the significance of the changes being brought by the Labour government's recent NHS White Papers. The focus will be on how incentives have changed and will change for those who in practice make most of the expenditure-driving decisions within the health service, namely the doctors. Because it is GP prescribing expenditure which is subject to the change in cash limit status, it is prescribing expenditure which is the main focus of the discussion in this paper.

Given the difficulty of keeping exactly to budgeted expenditure when much demand for health care not only requires immediate satisfaction but also cannot be forecast with 100 per cent accuracy, the result of the White Papers' changes may perversely be a progressive and significant weakening of control over total NHS spending. Many might see such an outcome as a good thing, meaning more money for the NHS, albeit at the expense of a greater burden on the taxpayer (either now or, if initially funded by increased government borrowing, later). However, unplanned NHS expenditure growth also has negative implications for the government's macroeconomic policies and for

achieving its desired balance of public spending priorities. Although a small percentage overspend on the NHS might appear to be insignificant in terms of the government's attempts to direct the overall economy (e.g. a one per cent overspend on the NHS in England would equate to a £350 million worsening of the Public Sector Borrowing Requirement (PSBR)), it would become more threatening to macro-economic stability if it were perceived as symptomatic of a lack of government will and/or ability to control public expenditure at all. Furthermore, overspending in the NHS would represent a divergence from the government's planned priorities as between health and other public expenditure programmes (social security, education, defence, and so on).

From a narrow health service viewpoint, any unplanned additional expenditure may not have gone where the NHS would, ex ante, most have liked any extra money to have been spent. The government may well respond to any overspending by seeking offsetting savings elsewhere within the NHS in a subsequent year.

2 CASH LIMITS IN THE NHS, TO 1998/99

For ease of presentation, the discussion in this paper is, unless otherwise stated, based on the specific structure of the NHS in England. However, where organisational differences in Northern Ireland, Scotland or Wales would affect the conclusions drawn, this is made clear and the implications of the differences are discussed. For example, there is no single cash limit applied to NHS spending throughout the UK. Rather, the NHS in England has a cash limit covering a majority of its total expenditure, the NHS in Scotland has a distinct cash limit applied to the corresponding portion of its total expenditure and so does the NHS in Wales. In Northern Ireland there is a cash limit applied to a part of the province's combined public expenditure on health and social services. However, for the purposes of this paper it is simpler, and not misleading, to discuss 'the NHS cash limit'.

2.1 Public Expenditure Control since 1948

Since the creation of the NHS in 1948, the government's chosen mechanisms for controlling public expenditure in the UK have changed. What has remained unchanged, however, is the fact that this choice has primarily been determined by two overall objectives and one major constraint. The perennial objectives are:

1. to help the government to fix taxation and plan public sector borrowing (or, conversely, public sector debt repayment), as major elements of its overall management of the UK economy;
2. to ensure that public expenditure priorities are achieved and not thwarted by over- or under-spending on some programmes.

The major constraint is the demand-led nature of expenditure in large parts of the government's programme, principally social security and health care. Of particular relevance to the discussion in this paper are the demand-led expenditures which require near-immediate satisfaction and so cannot be postponed to a later fiscal year: for example, patients' demands for prescribed medicines or emergency hospital care, rather than for elective surgery.

In the 1950s, public expenditure planning and control was by the 'Estimates' system. Every December, spending departments would

send estimates of their likely cash expenditure in the next financial year to the Treasury. These estimates were submitted under numerous sub-headings, around 2,000 of them across all government departments, corresponding to 'vote' headings, which Parliament would then vote to accept or reject. The NHS received its funding each year in this way until 1961. The main problems with the Estimates approach were that it was short-term, largely ignoring the future implications of current spending, and that it was vague about the volume of services and investment that the cash would buy. (Walshe, 1987, provides a concise history of changes in public expenditure control).

From 1961, in line with the recommendations of the Plowden Report (Cmnd 1432, *The control of public expenditure*, 1961), public spending plans were applied not only to the next year, but also looked five years ahead. The 2,000 vote headings were aggregated into a much smaller number of functional programme headings (such as transport or health). The plans were also expressed in volume, rather than cash, terms. That is, planned NHS and other public expenditure was in constant price terms, with the implication that whatever the rate of price inflation turned out to be, a corresponding amount of additional cash would be made available so as to fund the planned volume of expenditure.

The high rates of pay and price inflation experienced in the UK in the 1970s put paid to the volume approach. Large and unforeseen cost increases could no longer simply be funded unquestioningly. Besides which, the volume approach was seen to be stoking the inflationary spiral. High inflation led to higher public spending, which required increased public borrowing. This, according to the then prevailing macroeconomic orthodoxy, then contributed to higher inflation through the tendency for greater public borrowing to cause an increase in the money supply relative to the amount of real economic activity occurring.

In response to this, in 1976, the government introduced cash limits covering approximately 60 per cent of total public spending, but excluding areas seen as demand-led (Cmnd 6440, *Cash limits on public expenditure*, 1976). Among these non-cash limited areas were social security payments, European Community agricultural support payments, and what became known as the FHS elements of health expenditure. FHS expenditure consists predominantly of the costs of GPs (general medical services) and their prescribing, plus the costs of community pharmacy, dentistry and optician services. Under this system, to breach a cash limit required a specific case to be made to,

and accepted by, Parliament. However, any volume squeeze caused by unexpected price inflation would still operate only within the first year of the plan period, because in the next year the price inflation just experienced would be built into the starting point for the new plan.

In 1982 the then Conservative government reinforced the cash limit approach (Cmnd 8494, The government's expenditure plans 1982/83 to 1984/85, 1982). For the fiscal year 1982/83 and thereafter, supposedly rigid cash plans were set for each of the three years ahead. Although cash limits still only applied to about 60 per cent of total public expenditure, 100 per cent of it was to be planned in cash terms and this overall cash total was to be held to as a major part of the government's overall macroeconomic policy. The government presented its fundamental macroeconomic role as being to provide a stable fiscal and monetary environment into the medium term. 'Fine tuning' of the fiscal balance between tax revenues and public expenditures was out. Strict control of monetary conditions, including public borrowing, was in. The search for macroeconomic stability required plans to be stated clearly and then kept to. Unplanned public expenditure would be taken as a signal of a more general lack of government control over macroeconomic conditions and hence would be likely to undermine confidence in the British economy. This would, in turn, be expected to produce real and damaging economic consequences, such as higher interest rates and lower investment.

Keeping to cash limits was thus presented as a vital element of government economic policy. To cater for unplanned variations of expenditure while still trying to keep to a fixed overall public expenditure total, a larger contingency reserve was created. The government's stated intention was then not to be moved from the resulting total of planned public expenditure including the contingency reserve.

This approach to public expenditure planning and control remains, in essence, in place today. If anything the grip of expenditure planning is intended to be tightened. The current government's Comprehensive Spending Review, which reported in July 1998, announced 'firm three year plans' for each public expenditure programme. In principle, public expenditure was fixed not only for 1999/2000 but also for 2000/01 and 2001/02. It remains to be seen, however, how firm these plans will remain one year and two years after their first announcement, when the economic and political assumptions upon which they are explicitly based may look somewhat dated.

2.2 Controlling NHS Prescribing Expenditure

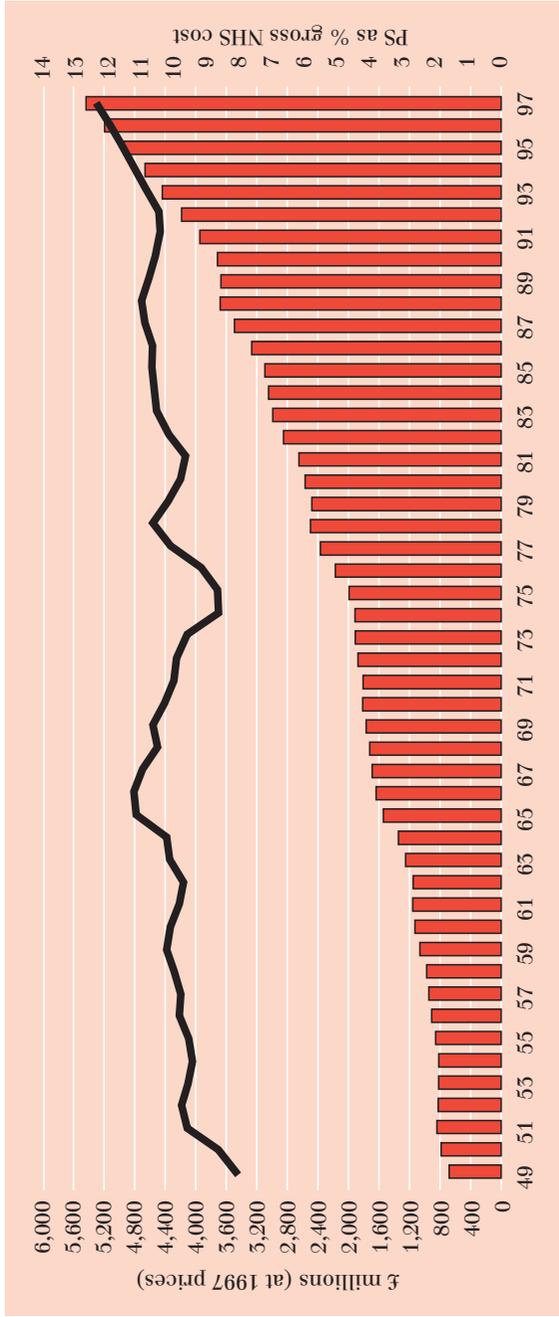
Webster, in his official history of the NHS, records a procession of government concerns with the seemingly inexorable growth of NHS expenditure, and within that particularly the growth in spending on prescribed medicines. The focus was primarily on the rate of growth of medicines expenditure, the large majority of which arises from prescribing by GPs. No analysis was attempted to determine what a desirable absolute level of medicines expenditure would be. Figure 1 shows that expenditure on pharmaceutical services (largely the cost of GP-prescribed medicines but also including community pharmacists' dispensing costs) outside hospitals has risen steadily in real terms, from £0.7 billion (1997 money) in 1949 to £5.4 billion in 1997. This expenditure has also risen, although less dramatically, as a percentage of total NHS costs: from eight per cent in 1949 to more than 12 per cent in 1997.

Discussion of the need for prescription charges started almost immediately after the NHS began and they were introduced in 1952, only four years into the life of the NHS, as part of an attempt to dampen pressure from GP prescribing. The Douglas (for Scotland) and Hinchcliffe (England) reviews of the causes of the rising medicines bill, both reporting in 1959, recommended education and exhortation of doctors to moderate their prescribing, but no coercion to do so. The current 'prescribing analysis and cost' (PACT) system introduced in 1988 to inform GPs of their prescribing patterns and costs relative to the average, may thus be seen as a natural descendent of the approach followed by UK governments since the 1950s. The overriding concern has been not to be seen to threaten doctors' right to prescribe what they judge to be necessary or, as Webster rather more starkly puts it, doctors' 'right to prescribe without limitation' (Webster, 1996, p.139).

The introduction of the voluntary GP fundholding scheme in 1991 could be seen as just one more step in a 43-year history of government attempts to encourage GPs to take more responsibility for NHS medicines expenditure. But it was a radical step. For the first time, 'overprescribing' GPs might experience financial loss as a consequence, although only in certain circumstances and only in the sense of foregoing the chance to spend extra money on their practices.

The GP fundholding scheme was created by the 1990 NHS and Community Care Act and the first GPs joined it in April 1991. GPs who became fundholders voluntarily agreed to the inclusion of their prescribing expenditure within an overall practice budget. This bud-

Figure 1 Gross* cost of pharmaceutical services (at 1997 prices) and as a percentage of total NHS costs, UK 1949–1997



Note: *Figures include dispensing fees, allowances, and prescription charges, at 1997 prices, as adjusted by the GDP deflator at factor cost.
Sources: The Government's Expenditure Plans (DoH).
 Health and Personal Social Services for England (DoH).
 Scottish Health Statistics. Health Statistics Wales.
 Annual Statistical Report (Northern Ireland CSA). Annual Abstract of Statistics (ONS).

get was also to pay for practice expenses, such as other practice staff and management costs, and the costs of buying a range of non-emergency, specialist hospital and community health care services for their patients. Increased expenditure on one part of the fundholding practice's annual budget was supposed to be matched by reduced expenditure on the others. But if the practice overspent the budget in total, it was the local health authority that was required to finance the deficit. No money would have to come out of the GPs' own pockets. The incentive for submitting to this scheme was that any surplus on the budget could in effect be retained by the practice (for up to four years) and invested in improving local health care provision. This included improvements to the GPs' own practice premises. Any deficit on the total budget would fall to the health authority, so for the GPs fundholding was a one-way bet.

This is not to say that the prescribing of other, non-fundholding, GPs has not also been the subject of expenditure constraints. But the incentives for non-fundholding GPs to limit prescribing expenditure are considerably weaker than for fundholders. All GPs have so-called 'target budgets' for prescribing (referred to as 'indicative prescribing amounts' prior to April 1994) but there are no financial penalties for GPs who exceed them, and only weak financial incentives for those who undershoot them. Some health authorities have attempted incentive schemes linked to target prescribing budgets. In some areas these take the form that if the non-fundholding practices in the district as a whole undershot their aggregated target prescribing budgets, then a portion of the undershoot would be reinvested within the district specifically in primary care (rather than simply being added to the large sums already being absorbed by hospitals). This provides only a very weak incentive from the perspective of any one practice, however. In other areas, the requirement that the aggregate district target be undershot has been relaxed, but the proportion of any 'savings' achieved by one practice which it is allowed to invest in service improvement in the practice is small (less than 50 per cent) and subject to a modest absolute upper limit per GP. Overall, as the Audit Commission (1994) put it, 'the Indicative Prescribing Scheme has been somewhat discredited to date as a means of controlling expenditure'.

The new White Papers' abolition of the GP fundholding scheme after eight years may therefore represent a noteworthy lessening of the pressure on prescribing. In expenditure control terms, the question is whether the simultaneous capture from April 1999 (in England,

Scotland and Wales) of all GP prescribing expenditure within the NHS cash limit for the first time can achieve its aims in the absence of strong incentives for prescribers. I return to this question in section 5.

Forty-nine years of expenditure control has kept the NHS cheap relative to health care systems internationally. According to OECD figures, per capita spend on health in the UK in 1996 was just 64 per cent of the average for all G7 nations taken together and was the lowest among those countries (i.e. Canada, France, Germany, Italy, Japan, UK and USA. Source: OECD health data 97: software for the comparison of 29 health systems, OECD, Paris 1997; quoted in Seitz et al., 1998). Estimated outturn NHS expenditure in England in the financial year 1997/98 is set out in Table 1, divided into the main categories relevant to the discussion of overall expenditure control. This shows that prior to implementation of the new NHS White Papers 83 per cent of total net NHS expenditure is subject to a cash limit; that is, all expenditure except that on the non-cash limited FHS. The largest single element of expenditure that is not cash limited is the remuneration paid to GPs, which totalled £2,208 million. The medicines prescribed by those GPs who were not fundholders in 1997/98 cost a further £1,920 million. The costs of dentists, pharmacists and opticians together accounted for £1,970 million. Prescription charges reduced

Table 1 Net expenditure on the NHS in England, current plus capital, 1997/98

	£ million
Hospital and community health services, cash limited	
family health services and related services	28,122
<i>Of which GP fundholders' medicines</i>	<i>2,204</i>
Non-cash limited family health services	5,776
<i>Of which non-fundholders' medicines</i>	<i>1,920</i>
Department of Health administration, central health and miscellaneous services (cash limited)	790
Total	34,688

Source: Department of Health, 1998. Figure 1.1 and Annex B, Figure B1. Estimated outturn as at April 1998.

the burden on the Exchequer by £522 million.

Although not advertised as such, the GP fundholding scheme increased the scope of the overall NHS cash limit. As more and more GP practices volunteered to join the scheme between 1991 and 1997, so the extent of the cash limit spread. By 1997/98 the medicines expenditure of fundholders equalled around six per cent of all NHS spending. That represents how far the fundholding scheme has extended the reach of the NHS cash limit since 1991: progressively and modestly, but significantly.

The total cash-limited sum of nearly £29 billion in England in 1997/98 is subdivided and allocated: both geographically to health authorities (and thence to fundholding GPs) and in terms of intended purpose, for example whether for capital or revenue expenditure. Each individual NHS agency - health authority, primary care practitioner, Trust - then makes its own budget allocations according to its own responsibilities and plans. But the cash limit itself strictly applies only to the total of all these expenditures and not separately to any subdivisions of it.

Over the years, the financial regime of the NHS has developed a large number of specific rules governing the circumstances where, and the extent to which, the money allocated at the beginning of the year to one agency (e.g. a health authority) or one purpose (e.g. maintenance of buildings and equipment) may in-year be redirected ('vired') to another. These rules are nowhere tighter than when governing the extent to which overspending in one year may be financed by advancing funding from the next. Thus, by midnight on 31 March the preceding year's expenditure is required to match very closely the total of funds that were set aside for that purpose at the beginning of the financial year¹. Aggregated to the national level, this ensures that the NHS cash limit holds in every year, and it is this rule which holds primacy. (A fuller description of the current NHS financial regime can be found in Prowle and Jones, 1997).

Thus, if a Trust overspends the amount agreed in its contracts with health authorities and other purchasers, then in the first instance one or more of those purchasers must try to find the funds by reducing expenditure elsewhere. Hence the adage among health authorities is that 'the purchaser always pays'. Up to 1996/97, the relevant Regional Office of the NHS Executive had the discretion to allow a Trust to over-

¹ Small overspends may be paid off in the following financial year; conversely, small sums resulting from underspends may be carried into the next year.

spend if the Regional Office could, in effect, find offsetting under-spends by other Trusts in its region. (Technically, this was achieved by allowing a Trust which fell short of its target rate of return to defer or cancel part of the dividend payment due from it on its public dividend capital, so that the Trust does not record a retained deficit as well as a too-low return on capital. But this was allowed only if other Trusts in the same region were able to pay additional dividends so that the region as a whole could meet its total target payment.) Since 1997/98, however, this discretion has been removed, making the NHS financial regime one notch more inflexible.

According to the NHS (England) Summarised Accounts for 1996/97, all 100 English health authorities kept within their cash limit for the year. This was despite the fact that 72 of the health authorities recorded a deficit in the year on an accruals basis. These deficits were funded by the use of working balances (i.e. chasing debtors and delaying payments to creditors) and by switching funds from capital to revenue. In 1997/98, for the first time, health authorities have been set income and expenditure targets as well as cash limits. From 1997/98 onwards all health authorities are supposed to break even on an accruals basis as well as keep within their cash allocation, except where there are 'deep-seated problems that cannot be resolved in a single year' (NHS (England) Summarised Accounts for 1996/97).

Non-cash limited expenditure is not subject to these formal constraints. Health authorities are nevertheless required to take measures to deter high spending even where there is no formal cash limit. But these measures boil down in practice to providing practice- and medicine-specific information on prescribing expenditures, and exhorting GPs not to spend more than their peers do. At the level of the whole district for which they are responsible, health authorities are also supposed to try and provide off-setting savings from their cash limited funds if non-cash limited expenditures look likely to exceed planned amounts. However, the scope for such off-setting savings for most health authorities, in an environment where funding levels have been persistently tight in the face of continuous 'cost pressures', is small. Thus, higher than planned expenditure on non-cash limited areas is likely to mean higher than planned NHS expenditure in total. This issue is discussed further in the next section.

3. THE 1997/98 WHITE PAPERS – EXTENDING THE NHS CASH LIMIT

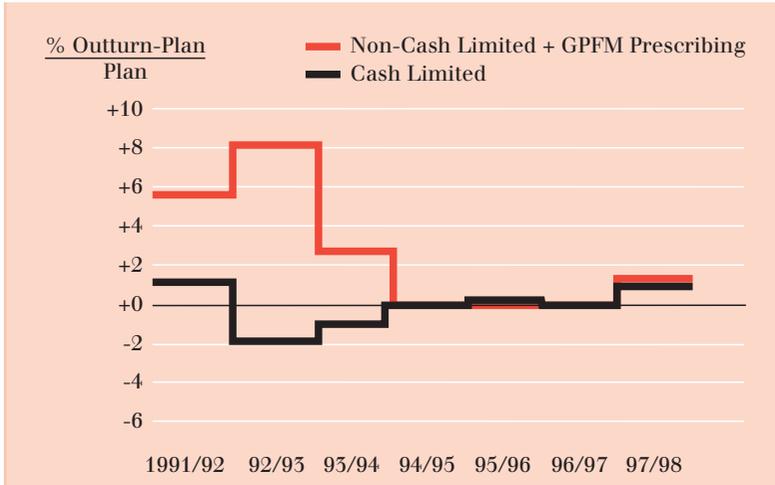
Labour's NHS White Papers' extension of the cash limit to include non-fundholding GPs' prescribing will bring within that limit a further six per cent of total NHS expenditure. This will extend the scope of the cash limit in one step by an amount similar to that which the GP fundholding scheme has brought progressively within the limit over the seven years since April 1991. The reasoning behind this change may be along the lines that: 'half the GPs have already volunteered to be cash limited and that has not caused unbearable difficulties, so now let's cash limit the other half too'. Thus, based on the 1997/98 figures shown in Table 1, around 89 per cent of total NHS expenditure will come within the overall cash limit from 1999/2000 onwards, up from 83 per cent now. This proposed change does not mean that there will be a total national budget for medicines expenditure alone. It does mean that higher medicines expenditure should, in future, be balanced by lower expenditure under any or all of those non-medicines headings which fall within the definition of the overall NHS cash limit.

The government's desire to maintain the integrity of the NHS cash limit after April 1999 is apparently as great as it has been hitherto. The NHS financial regime, with its pressure for health authorities and Trusts to meet their financial targets spot on at midnight on 31 March every year, and its rules to try and prevent generalised breaching of those targets, remains just as strict as before. However, the future inclusion of the total GP medicines bill within the cash limit is likely, other things being equal, to make that cash limit harder to keep to, because prescribing expenditure is demand-led and difficult to predict exactly. That is why, when the NHS cash limit was first created in 1976, the GP medicines bill was left outside it. Variations in demand for specialist, non-emergency diagnosis, treatment and care by the NHS are managed, and the available services are rationed, by waiting lists. But there are no waiting lists for the receipt of medicines.

It is difficult to quantify from published data the degree of unpredictability of the GP medicines bill. Indeed, as one expert on government accounting (who subsequently became a senior Treasury civil servant) put it:

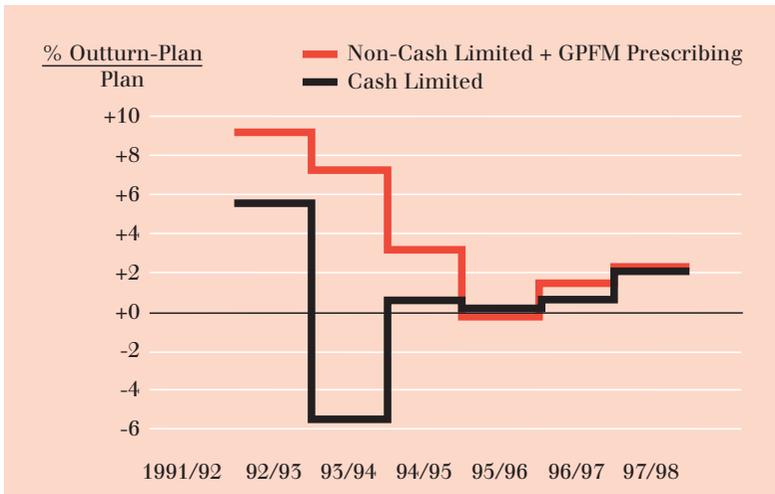
'In part, the continued credibility of the public expenditure control system may be due to the fact that the data is presented in such a way as to make it difficult to assess whether public expenditure has been

Figure 2 Deviation of outturn expenditure from that planned one year earlier, expressed as percent of planned expenditure – NHS England



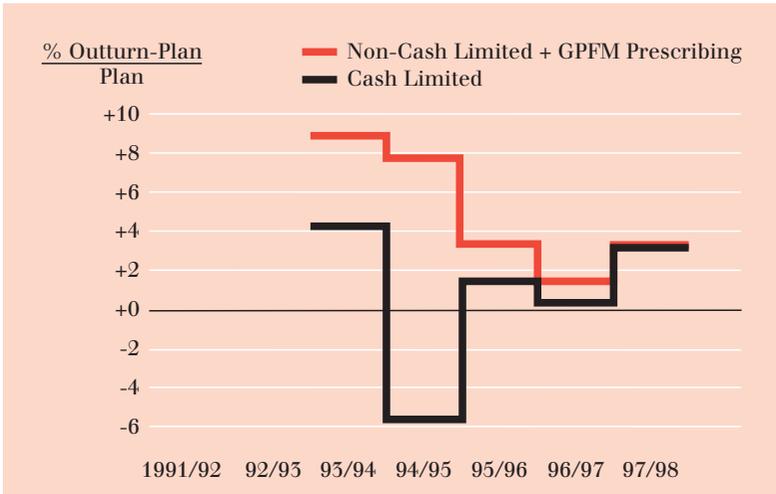
Source: Government Expenditure Plans, published annually 1991 to 1998.

Figure 3 Deviation of outturn expenditure from that planned two years earlier, expressed as percent of planned expenditure – NHS England



Source: Government Expenditure Plans, published annually 1991 to 1998.

Figure 4 Deviation of outturn expenditure from that planned three years earlier, expressed as percent of planned expenditure – NHS England



Source: Government Expenditure Plans, published annually 1991 to 1998.

contained or not' (Likierman, 1986)

However, in Figures 2 to 4 I have attempted to provide a broad indication of how predictable prescribing and other non-cash limited NHS expenditure has proved to be at an aggregate level. The charts are drawn from 'The Government's Expenditure Plans'² for the NHS in England, as published in each of the last seven financial years. They compare the outturn level of expenditure in each year with the level of expenditure that had been planned for that period, respectively one, two and three years earlier. The comparison of plan with outturn is shown for cash limited and non-cash limited expenditure separately, but with GP fundholders' prescribing costs added to the non-cash limited total. The reason for this less than ideal presentation of information lies in the way that the government's expenditure plans are presented in published sources. Specifically, data for planned expenditure on non-cash limited medicines separately are only available from the latest three years' 'Government expenditure plans' and not

² Although published alongside plans for cash-limited expenditure and under the common title 'The Government's Expenditure Plans', future spends on non-cash limited programmes are necessarily actually projections of expected outturns rather than 'plans'.

earlier. Even then, this planned sum has only been shown for one year ahead: plans for the second and third future years are still not shown to this degree of disaggregation.

Figure 2 compares outturn expenditures with the amounts that had been planned just one year earlier. For example, Department of Health (March 1997) shows planned NHS cash limited expenditure for the next financial year, 1997/98, to be £25,685 million. The outturn was £25,919 (latest available estimate, published in April 1998, source: Department of Health, 1998). Thus the outturn level of cash-limited expenditure was £236 million (0.92 per cent) over plan. The picture for non-cash limited expenditure plus GP fundholders' prescribing costs was a 1997/98 outturn (estimated) of £7,980 million, compared with the plan of £7,875 made one year earlier. That is, the outturn non-cash limited expenditure was £107 million (1.36 per cent) above the level that had been planned one year previously.

Figure 2 shows that since 1994/95, when GP fundholding was already fairly widespread (covering 35 per cent of the English population), year-ahead plans for NHS expenditure in England have been pretty well achieved, whether cash-limited or not. In the three years before 1994/95, however, outturn non-cash limited expenditure plus GP fundholders' prescribing was repeatedly well over (three to eight per cent over) the sums planned one year earlier. Figure 2 also shows that, in those earlier three years, planning of non-cash limited expenditure plus fundholders' prescribing was considerably less successful than for cash-limited expenditure. This was a continuation of a long-established pattern. As Webster (1998) puts it:

'During the 1970s the government's armoury of expenditure controls and management reforms had been effective in bringing the HCHS [Hospital and Community Health Services, cash limited] under a much stronger regime of containment. Any satisfaction at this achievement was offset by irritation and embarrassment over the failure to achieve anything like the same degree of control over the independent-contractor services. To add to the indignity of this situation, owing to forecasting errors, the FPS [Family Practitioner Services, now referred to as Family Health Services or FHS, and non-cash limited] regularly exceeded their estimates.'

Over time, the total of non-cash limited expenditure and all GPs' prescribing costs appears to have become more predictable, or at least better predicted. The question is whether this was primarily due to the spread of GP fundholding, which meant that more and more of the

medicines bill effectively came to be cash limited and, if that was the reason, whether this predictability will deteriorate once fundholding is abolished in April 1999.

Figure 2 also appears to show a gradual but steady rise after 1992/93 in outturn cash-limited expenditure relative to what was planned one year earlier. In 1992/93 outturn cash-limited NHS expenditure in England was 1.93 per cent below plan; by 1997/98 the outturn was 0.92 per cent over plan. These numbers may seem small, but a variation of 2.85 percentage points (0.92+1.93) would be equivalent to £1 billion in the NHS in England today.

Figures 3 and 4 show that plans made two and three years earlier are, unsurprisingly, kept to rather less closely than those made just one year ahead. These two Figures also reinforce the impression that outturn has been held closer to plan over time. (The data for the deviations of outturn from planned expenditures, from which Figures 2-4 have been drawn, are given in the Appendix to this paper.)

While the rules of the NHS financial system will be no more flexible than before, the unpredictability of the expenditure which is required to be kept within the cash limit will increase after 1 April 1999, when it will include all GP prescribing for the first time. Keeping the lid on will become harder. I do not wish to overemphasise the case, though. The cash limit is being extended by roughly £2 billion out of total NHS expenditure of £35 billion (England, 1997/98 figures) and unplanned variances of expenditure in the recent past have only been around one per cent of the total non-cash limited budget planned a year before. However, although it is modest, this increase in the pressure on the cash limit which will result from stretching it to cover a larger proportion of NHS spending, coincides with an apparent weakening of the means available to counteract that pressure.

This weakening follows from the proposed White Paper changes in the incentives facing GPs to control their prescribing and other expenditures. The White Paper changes themselves are discussed in the next section, and their probable impact on incentives in section 5. Added to that, there is the risk that if the weakening of budget incentives for GPs leads to a noticeable increase in overspending of budgets all around the NHS, then the impact might become cumulative: the more people bust their budgets, the greater will be the incentive for others to do the same. This cumulative risk is discussed in section 6.

4 THE 1997/98 WHITE PAPERS – ABOLISHING GP FUNDHOLDING

The Labour government's 1997/98 NHS White Papers are unanimous in bringing the GP fundholding scheme to an end. 1998/99 is to be the last year of that scheme in England, Scotland and Wales; 1999/2000 in Northern Ireland. In its place will be put a variety of structures, with a different form and nomenclature in each country of the UK. What all of these structures will have in common is that, although GPs are intended to have strong influence over local health care expenditure decisions, individual practices will not have their own cash budgets and will not be able to keep within the practice any savings they achieve. Where there are hard, cash budgets in future, these are to be held collectively by large groups of practices, covering around 50 GPs, rather than by individual practices.

The planned arrangements vary between countries. In **England**, 'Primary Care Groups comprising all GPs in an area together with community nurses will take responsibility for commissioning services for the local community' (Secretary of State for Health, 1997, paragraph 3.18). These Primary Care Groups cover 'natural communities' with average populations of around 100,000, meaning that there are around five Groups within each current health authority area. There are, typically, around 50-60 GPs from 20 or so different practices in each Primary Care Group. GPs have not had the option of choosing which Primary Care Group to join: there is only the one possibility, determined by location. Each Group will remain accountable to its local health authority and can agree to take any one of four increasing levels of responsibility. These range from simply acting as adviser to the health authority which continues to commission health care; to the Group having its own budget for buying all health care for the population it serves³; or even to becoming a free-standing Primary Care Trust with the added responsibility of providing local community health services as well as buying all types of services.

Where a Primary Care Group does take on the health care budget for its population, this will be held at the Group, not the individual GP practice, level. Further, a Group may overspend its budget, apparently without penalty. As the Minister of State for Health, Alan Milburn,

³ With the exception that the Primary Care Group or Primary Care Trust budget will not cover GPs' own remuneration, nor that of community dentists, pharmacists and opticians.

put it when speaking during Parliamentary Questions in the House of Commons on 28 April 1998, 'I say to him (the questioner), to the House and to GPs throughout the country that no GP will run out of cash, that patients will be guaranteed the drugs and the treatment that they need and that if a Primary Care Group overspends, the overspend will be catered for within the health authority's general allocation' (source: Hansard). It is the health authority that will remain responsible to the Secretary of State for balancing the combined budgets of all health care commissioners in its area: responsibility without power, an unenviable combination.

In **Wales**, the planned arrangements are similar to those for England but not exactly the same (Secretary of State for Wales, 1998). In Wales, the GP-centred commissioning groups are referred to as Local Health Groups and explicitly include not only GPs and community nurses but also dentists, pharmacists, and (non-NHS) social services staff and representatives of voluntary organisations. The boundaries of the Local Health Groups coincide with the 22 Welsh unitary local authorities – the bodies responsible for providing personal social services. This means that the average population coverage of Welsh Local Health Groups is around 130,000. The health care commissioning and budget-holding options for these Groups are the same as for the English Primary Care Groups⁴. Thus budgets will be held at Group level (or by the host health authority), not by individual GP practices. Again it is the health authority – which typically has four or five Local Health Groups within its area – and not the GP-centred Group who will be financially accountable to the Secretary of State (for Wales in this case, rather than the Secretary of State for Health).

The plan for **Scotland** has a number of rather different features but, nevertheless, in common with the rest of the UK, removes direct budget incentives from individual GP practices (Secretary of State for Scotland, 1997). GPs in Scotland are to be grouped together in Local Health Care Co-operatives but these groups appear unlikely to be given control of budgets, not even collectively. Instead, the Co-operatives are to combine with community and mental health Trusts into what are to be called Primary Care Trusts, to provide (rather than commission) all primary, community and mental health care for the local population. The Scottish health boards (equivalent to English

⁴ Although social services representatives are involved in the decision making within Welsh Local Health Groups, there is no proposal at this stage to increase the scope of those Groups' budgets to include social services in addition to health care.

and Welsh health authorities) are, in effect, to commission health care services from both Primary Care Trusts and acute hospital Trusts; although the word 'commission' does not appear in the Scottish White Paper, let alone 'purchase' or 'buy', in the context of health care services. Each of the 15 health boards will have just one Primary Care Trust and (other than in Glasgow) one acute Trust within its boundaries. Thus each Primary Care Trust will cover, on average, a population of 350,000 (although the sizes of individual Trusts will vary widely). As currently described, there will be no practice-level budget-related incentives for GPs in Scotland after 1 April 1999.

The arrangements for **Northern Ireland** are not yet fixed, but the government has published a consultation document (Secretary of State for Northern Ireland, 1998) which indicates the range of options for the eventual outcome. Common to these is the abolition of GP fundholding, although one year later than elsewhere in the UK. One option proposed would, in essence replicate Welsh-style groups of GPs, community nurses, other health professionals and social workers. These groups would hold budgets for all health and personal social services in their area. Under one possible option they would be 'Primary Care Groups' covering populations of 50,000-100,000 and would be ultimately accountable to the host health and social services board (of which there are already four). Under an alternative option, 'Primary Care Partnerships' of GPs and other primary care professionals, covering populations of 25,000-50,000 would work with 'Local Care Agencies', which would combine the functions of the current health and social services boards and the provider Trusts and serve populations of around 200,000-300,000. This second option is thus more akin to the Scottish approach.

Once these reforms are in place throughout the UK, no individual GP practice will be able to benefit financially from savings it may achieve in the costs of the care it provides or commissions. (The nature of the existing incentives for GPs and what will happen to them when Labour's White Papers are implemented, are discussed in section 5 below). This has the advantage that it removes the apparent risk to the GP-patient relationship of the patient having cause to fear that the treatment they are offered may be affected not only by their needs but also by the financial interests of the doctor. It has the disadvantage, however, of removing a pressure on those GPs who are currently fundholders to keep health care costs in aggregate under control.

5 INCENTIVES FOR GPs

The GP fundholding scheme was voluntary and has provided individual GP practices with a clear financial incentive to participate. Savings achieved against the practice's total budget could be reinvested at that practice's discretion in a range of ways aimed at benefiting the health care of the local population. Among these has been the option of improving the quality of the practice's own premises, and the standard of equipment in them, thus enhancing the value of assets that the practice's partners may own. Thus a practice spending below its total fundholding budget could turn those savings into something of benefit to their patients and possibly to their own balance sheets. Budget overspends, on the other hand, do not have to be funded by the practice but instead are covered by the local health authority.

To ensure the success of the fundholding scheme – i.e. substantial uptake and few subsequent withdrawals from the scheme by GP practices failing to keep within their budgets – budgets were set at fairly generous levels (see, for example, Smith, 1997, p.10). Hence GP practices could be reasonably confident that measures to restrain their costs would produce savings against their budget and hence financial benefit for them. If GPs considered their budgets to be so tight that there was a significant chance of overspending, that would weaken or remove the incentive to try to achieve savings at all. This is because if the result of such efforts was expected merely to be a smaller overspend against budget than otherwise would have occurred (rather than a bigger underspend), then there would be no expected 'savings' for the practice to invest at its discretion. In a sample survey of English fundholders in 1993/94 (when fundholders were already covering 41 per cent of patients in England and Wales), the Audit Commission found that three quarters of them made a saving while just a quarter overspent (Audit Commission, 1995). On balance, fundholders in England had underspent by £64 million, 3.5 per cent of the total budgets allotted.

Even if budgets had not been so generous, they would in many cases have been relatively easy to keep within for the reason that they were largely based on historic expenditures by the practices concerned. That is, relatively high-spending practices were allocated correspondingly greater budgets than historically lower-spending practices. Because medicines expenditure had not previously been constrained, 'there was certainly no reason why a practice would want to spend more on drugs per se after fundholding since it could pre-

scribe all it chose without cost before the Reforms' (West, 1997, p.84).

In a review of the literature, Baines et al. (1997) found a clear consensus that fundholding practices generally achieved one-off reductions in their prescribing costs relative to those of non-fundholding practices. The prescribing costs of fundholders did not actually fall but, in the first year or two after becoming fundholders, the growth in their prescribing expenditures was slower than that of non-fundholders. Subsequently the rates of growth of both fundholders' and non-fundholders' prescribing expenditures were similar, with fundholders' expenditure remaining below the level that would have been expected in the absence of fundholding. The main mechanisms for achieving this relative saving appear to have been: increased generic prescribing, limitations on prescription volume and the use of practice formularies. It is less clear whether fundholding had any impact on referral rates for elective surgery and other specialist services.

Fundholding GPs thus appear to have behaved in a more price-sensitive way when prescribing than have their non-fundholding colleagues. Pharmaceutical companies appear to have responded to this by changing the relative prices at which they have launched new medicines in the UK. Towse and Leighton (1998) have studied the UK prices of new follower medicines, i.e. those new medicines that are launched in therapeutic areas where at least one alternative medicine already exists. They have found that whereas in the past follower medicines would normally be priced above the price of the existing market leader, this is no longer so. In the 1990s, the UK prices of new follower medicines have usually been set at a discount to the existing market leader's price, presumably in an effort to win market share with price-sensitive prescribers.

Two key characteristics of the GP fundholding scheme were thus:

1. that it was voluntary. No GP practice whose partners did not wish to do so had to take responsibility for the commissioning of health care services or keeping to a fixed budget;

2. that participants who made savings could expect to be able to reinvest those savings themselves, in ways which would directly improve the welfare of their own patients and/or which would increase the value of the practice assets (premises and equipment) they owned.

Both of these features are removed by the new White Papers' proposals. Membership of Primary Care Groups or their equivalents is

compulsory and, furthermore, the group is chosen *for* the GP practice not *by* it. GPs who resisted the lure of fundholding are grouped with others who embraced it enthusiastically. All GPs will be able to retain their much cherished independent contractor status. Yet they will be expected to reach joint decisions and respect (at least in some English and Welsh cases) a common budget ceiling with 50 or more other GPs, some of whom will be friends and colleagues but some who may be long-standing rivals. The first danger from these White Paper proposals is, therefore, that GPs who chose not to volunteer to submit to budget disciplines under the fundholding scheme, may not be inclined to keep to budgets under any circumstances.

The greater weakness, however, which will apply for all GPs, whether formerly fundholders or not, is the lack of direct incentives in the White Papers' world, as currently specified, for individual practices to be economical in their use of health care services. Without practice-level incentives, a GP practice in an average-sized Primary Care Group containing 15-20 practices might reasonably suppose that:

1. any saving it might achieve would be likely to be swallowed up by overspending among one or more of the other 14-19 practices over which it can exert no effective control; and
2. if it did achieve a saving that was not cancelled out elsewhere within the Group, it would still only stand to receive between one fifteenth and one twentieth of that saving itself.

This hardly represents a strong motivation for economy.

The English White Paper states that 'It will be open to the (Primary Care) Group to agree practice-level incentive arrangements ...' (para. 5.19). However, hard practice-level budgets would represent a clear return to, officially reviled and rejected, GP fundholding. Although one major objection to fundholding – that it caused inequity of resource allocation between different groups of patients – is removed by placing all GP practices in Primary Care Groups (or their equivalents), practice-level incentives must give some discretion to GPs if they are to be effective. Individual practices with their own budgets might want to purchase different packages of health care services, which could hinder moves to greater equality of access for all patients in a Primary Care Group's area. Practice budgets could also entail a greatly increased number of contractual arrangements between care providers and care purchasers, undermining attempts to cut bureaucracy. It is therefore unclear whether effective practice-level incen-

tives can be created within Primary Care Groups. If they cannot be created, the threat is clear: without the prospect of receiving clear benefits at the practice level, GPs may quite reasonably see keeping to a budget as pain without gain.

Baines et al. (1997) compared the experience of fundholding GPs with that of non-fundholders, in respect of their prescribing expenditures. For the latter group the only incentive for keeping within their indicative prescribing amounts ('target prescribing budgets') and not making undue referrals to specialist services, was the satisfaction of being seen to be socially responsible in the use of resources within their host health authority's area. Baines and colleagues found that 'The fundholders' potential for earning real and usable financial surpluses appears to have exerted a stronger and more rapid influence on behaviour than did the non-fundholders' prospects of generating purely nominal surpluses' (p.19)

In an earlier study comparing fundholders with non-fundholders, Glennerster et al. (1994) found that practices who were not fundholding considered the indicative prescribing scheme alone as providing them with no incentives. The inducement for non-fundholders to prescribe less in return for the vague benefit of primary care in the area generally, was found to have no impact.

Other authors have taken advantage of the wide range of GP-based commissioning models that have arisen in the NHS during the 1990s, to try and discern the impacts of different types of incentives. Smith et al. (1998) conducted a national survey of commissioning models found in English health authorities. They classified the models in use around the country under six headings:

1. GP multifunds – where large groups of fundholding GPs have voluntarily pooled their practices' management allowances to finance a common management/administration which provides services such as finance and personnel management to multifund members and represents them jointly in negotiations with local health care providers and with the health authority;

2. fundholding consortia – of practices negotiating jointly for specific services;

3. locality commissioning – where health authorities divide their population geographically into sub-areas and for each sub-area establish GP fora with identified lead GPs and dedicated health authority managers liaising with them;

4. commissioning based on groups of practices – similar to the previous type of commissioning but based on practices rather than populations within defined geographical boundaries;

5. authority-wide GP group – i.e. just one GP forum for the whole health authority area;

6. a catch-all category of any other commissioning arrangements.

Among their numerous findings was that, in the view of senior health authority managers, GP multifunds had the clearest impacts on prescribing and referral practice of any of the six types of commissioning models. In multifunds, although GPs pool their management allowances and share services, any savings they achieve against their practice budgets are specific to the individual practices and do not have to be shared with the other practices in the multifund. Conversely, however, locality commissioning groups were most frequently cited by health authority managers as having had a beneficial impact on service quality; whereas multifunds were least often cited.

The weakness of group-based incentives relative to individual practice-based incentives has also been shown by the King's Fund's national evaluation of Total Purchasing Pilot Projects (TPPs). These projects were set up to investigate the feasibility and impact of extending the scope of fundholding GPs' budgets beyond payments for practice staff, prescribed medicines, and a range of elective hospital and specialist community referral services. Bevan et al. (1998) report a number of pertinent findings from their survey of the 53 first wave TPPs, including that multi-practice TPPs were more likely than single-practice TPPs to show variations in actual to planned spending and found it more difficult to manage expenditure within budgets. More specifically, for the 37 TPPs who answered the relevant questions, ten out of the 26 multi-practice TPPs, but only one of the 11 single-practice TPPs, had experienced 'any kind of financial difficulty/crisis'.

Bevan et al. also identified what they considered to be an important third type of GP apart from fundholders or non-fundholders, namely: 'non-fundholding GPs in fundholding practices'. These were characterised as GPs who have been happy to accept the benefits of fundholding for their patients – e.g. practice-based services - provided that fundholding did not interfere with their clinical autonomy (Bevan et al., 1998). Abolishing fundholding, while at the same time stressing the leading role of GPs in determining how the NHS's resources are spent, risks putting all GPs into such a category: happy to share any gains their Primary Care Group (or equivalent) obtains but unwilling

to bear their share of any pain involved.

The NHS Confederation has published a report of a series of 17 workshops held around the UK during February-April 1998 to discuss the proposed Primary Care Groups and their Scottish and Welsh equivalents. The workshops were attended by a total of 3,000 participants: NHS managers, GPs and their practice staff, other clinicians, community nurses and a range of other health care professionals. Among the main findings from these workshops was that 'With no guarantee of conformity [between GPs, and others, within a Primary Care Group or equivalent] rooted in like-mindedness, no sanctions against independent contractors, and no financial incentives to participate there is understandable concern about how effective accountability is to be achieved'. (Marks and Hunter, 1998, p.41)

6 THE RESULT: CALLING THE BUDGET BLUFF?

The reasons why UK governments past and present have not attempted to cash limit all areas of public expenditure are linked to the demand-led, unpredictable nature of some expenditures. If there had been no perceived danger in imposing a cash limit on social security spending and several billion pounds of NHS expenditure each year, then it is a safe bet that cash limits would have been set in 1976 or 1982. Cash limits appeal to governments both because of their expected macroeconomic benefits and because they enable the government to achieve its priorities between the various public spending programmes.

The decision to bring all GPs' prescribing costs under the total NHS cash limit, as proposed in Labour's recent NHS White Papers, presumably rests, therefore, on a belief either that the demand for prescription medicines can now be managed (controlled), or that the Department of Health's ability to predict the demand for them across the country as a whole has improved, or both.

As discussed in the previous section, there are reasons to fear that implementation of the recent NHS White Papers will weaken the incentives for GPs to keep within budgeted expenditure levels unless a mechanism can be found to give GPs effective, individual practice-based, financial incentives. Those GPs who have been fundholders will otherwise lose their direct incentives to do so when that status is removed from them. Furthermore, they will be forced to share their Primary Care Group's (or equivalent's) budget with other practices, including those GPs who in the past chose not to become fundholders (and even small practices have had the option to become fundholders, by sharing and co-managing a budget with another practice or practices). The belief that GPs' prescribing and other expenditures will generally be more manageable in future seems, therefore, in the absence of practice-based incentives, to be an act of faith.

While the incentives for the over 50 per cent of GPs who were fundholders to economise risk being significantly weakened, the underlying difficulty of predicting patient needs at the level of accuracy needed to hit annual budgets without leaving large unspent contingency reserves, remains. In the face of this, if some GPs within a Primary Care Group are seen to be flouting budget constraints, there will then be a strong temptation for other members of that Group to do

likewise. If some Groups are seen to be disregarding their budgets, other Groups will be tempted to emulate them. If this forces some health authorities to overspend, despite the disapproval of the Department of Health and the Treasury, this may increase the willingness of other health authorities to risk the same disapproval.

Patient pressure may reinforce the temptation. A well-informed patient, aware from the media of unpenalised overspending by GPs, Primary Care Groups or health authorities in other parts of the country, will be unimpressed by what they perceive as any attempt by their own GP to deny them a prescription or access to any other health care service (however medically justifiable that denial may be).

Furthermore, despite apparent initial enthusiasm for the new NHS White Papers, representatives of the medical profession subsequently changed their minds and voiced strong opposition to Primary Care Groups. Despite the statement by the Minister of State for Health, Alan Milburn, to the House of Commons on 28 April 1998 that no GP would be allowed to run out of cash (which was quoted in section 4 above), the British Medical Association was not mollified. As reported in the national press in May 1998, Dr Chisholm, the British Medical Association's chief GP representative, warned the Minister of Health that 'GPs were very angry, and growing increasingly worried about the proposals to make them join Primary Care Groups'. In response to this warning, the Minister, Mr Milburn, 'reassured the GPs that they will retain their status as independent contractors in the NHS, they will keep their clinical freedom and they will be allowed to overspend on their annual budgets, in spite of cash limits' (both quotes taken from an article in *The Independent* of 16 May 1998 headed 'Minister holds talks to avert revolt by GPs', emphasis added).

As a result of subsequent negotiations between the General Medical Services Council (GMSC) and the Department of Health, GPs have also won protection for that part of the total budget their Primary Care Groups will receive which is for practice infrastructure and management. A Primary Care Group that overspends its budget for hospital referrals, or for prescribing, will not be called upon to pay for that overspend by cutting back on the practice expenses of its member GPs.

An overspending GP cannot be sacked, after all. A government or health authority which tried to terminate the contracts of overspending GPs would in effect be cutting front line health care services and directly hurting those GPs' patients. But budgets act to restrain expenditure only if those who determine expenditure perceive them to be

genuinely fixed. Once budgets may evidently be broken with impunity, they can no longer achieve their purpose.

It is upon health authority managers, not GPs, that the pressure of ministers and of the central management structure of the NHS can be brought to bear most strongly. The chief executive of an overspending health authority could, in principle, soon become an ex-chief executive. By extension, the same might be said of a health authority director of finance, or indeed any other senior health authority manager. Whether from a sense of duty or from fear of curtailed careers, health authority managers may therefore attempt to keep the lid on the budget in their district, despite the actions of individual Primary Care Groups or individual GPs. There is a range of measures that they can adopt in attempting to do so, but many of them are purely temporary and the others are painful and highly unpopular with the public.

Temporary measures which might be turned to 'in-year' if and when likely overspending problems become apparent, whatever the cause of the overspend, are:

1. Health authorities may broker resources between those Primary Care Groups (and their Welsh, Scottish and Northern Irish counterparts) in their districts that overspend and those that underspend. Regional Offices of the NHS Executive may do the same between health authorities. Brokerage can ultimately only work, of course, as long as the NHS in total has not overspent its cash-limited funds; i.e. as long as there are underspenders to bail out the overspenders.

2. Trusts may be required to delay, diminish or cancel planned expenditures on one-off projects or new service developments. For example, raiding the block capital funds allocated to Trusts to maintain or repair buildings and replace worn-out equipment, in order to find funds to meet an overspend on the revenue account, is common. But this can only ever be a very short-term expedient and is one which inevitably stores up future problems: deferred repairs become more expensive; worn-out equipment is unsafe; delay to new services means lost patient benefits and damage to staff (and community) morale.

3. At the year end, when some budgets inevitably remain unbalanced, there remains a very short-term measure to cover the cracks, namely very short-term borrowing from non-NHS sources on 31 March and repayment one day later, on 1 April, in the new financial year. But this dubious practice only carries the overspend forward into the next year: a problem delayed is not a problem solved.

Less temporary solutions to over-running budgets and to generalised budget-breaking mean cutting back other parts of the health care budget. Thus, if medicines expenditure, or any other area of health care spending, were to prove unmanageable and no additional Exchequer funds were forthcoming to meet the unplanned costs, then other health services would eventually have to be cut. Health authorities have long and bitter experience of making unpopular service cuts in one part of the NHS cash-limited budget in order to fund over-spends in another part. For example, every winter some health authorities find themselves with unexpectedly high and/or costly levels of demand for emergency hospital care which threaten to break their budgets. Every year, the press records the 'cuts' proposed and implemented in response, and the horrified public reaction to them. Bringing all GPs' prescribing costs within health authorities' cash limited budgets without strong incentives for GPs to conform to those budgets, can only make such unplanned and unpopular service cutting a more frequent event.

A trawl through the last two winters' editions of the Health Service Journal reveals that a range of sacrifices may be offered to balance budgets. Emergency services cannot be cut and, in recent years at least, mental health services have also tended to be protected. Any other area of health services may suffer, but by far the most common sacrifice has in practice been to cut back on elective surgery. The result of this is, of course, cancelled operations; greater delays for patients, who must endure pain and disability for longer before receiving treatment; and growing waiting lists, which embarrass government ministers. During the winters of 1997/98 (which was exceptionally mild) and 1996/97 (a pre-election period), the Health Service Journal included news stories on the following examples of health authorities' making unplanned service cuts in-year in order to try and keep within that year's budget:

- Barnet – closure of an elective surgical ward, delay to routine elective surgery, restrictions on non-urgent outpatient appointments;
- South Essex – reduced community services and non-emergency patient transport, and less support to services provided by voluntary groups;
- Cambridge and Huntingdon – abolition of school nursing service and reduction of health visiting;
- Northamptonshire – cut in health visiting;

- Bedfordshire – reduced health promotion expenditure, plus cuts at all provider trusts;
- East London and the City – increased waiting times for elective surgery;
- Oxfordshire – delaying non-urgent cardiac surgery;
- Merton, Sutton and Wandsworth – limiting non-life threatening operations, principally for eye, hip or wisdom tooth surgery.

There will be many other examples, from all around the UK, which were not specifically named in the pages of the Health Service Journal. For example, a survey by that Journal, reported in its 27 November 1997 edition, found that ‘64 per cent of trusts anticipate cutting services and 18 per cent think they may be forced to close wards or units because of their projected overspends’. Of course, these overspends arise for many reasons, many of them unconnected with the referral and prescribing practices of GPs. The point is, however, that increasing the pressure on budgets from any direction can ultimately have the kind of undesirable repercussions listed above or, alternatively, can increase the frequency with which budgets are breached because of the pain involved in holding to them.

Given the unpopularity of cuts such as those just listed, it is remarkable how little budget breaking appears actually to have gone on hitherto. Figure 2 earlier showed how close, in aggregate, outturn NHS expenditure in cash limited areas has been to what was planned at the beginning of the year, throughout the last seven years. In general, it seems that aggregate budgets have held. But what of the future? There is already a hint in the tentative trend shown in Figure 2 for the outturn total of cash limited NHS expenditure increasingly to exceed the planned sum.

The above-trend funding increase of £3.1 billion awarded to the NHS in England for 1999/2000 that was announced in July 1998 as a result of the government’s Comprehensive Spending Review, may create some initial headroom and so ease the pressure on NHS budgets in that year. However, much of this increase has already been pre-empted to fund reductions in waiting lists and the implementation of the NHS’ new IT strategy. Furthermore planned increases in NHS funding in subsequent years are less generous. By then the ever-growing demands on the NHS by its patients and the pressure for better pay by its staff are likely to make budgets extremely tight once more.

A reason why breaches of NHS budgets may be temporarily sup-

pressed in 1998/99 is that the pressure on health authority managers to try and keep within budget is currently stronger than ever, because the number of health authorities (and Trusts) is to be reduced through mergers and the number of senior managerial posts will fall correspondingly. Competition for the reduced number of jobs available is likely to be strong. Convicted overspenders need not apply. In 1999/2000 and thereafter, however, when any merger wave has passed by, the credibility of the threat of dismissal may recede correspondingly. Coinciding with that, the White Paper proposals will bring all prescribing within the national cash limit and simultaneously weaken the incentive for GPs to control the cost consequences of their prescribing and referrals unless effective practice-based incentives can be devised. Keeping within budget will clearly become harder and this will be true in every health authority simultaneously.

The clear danger therefore exists that at some point the existing steady stream of overspending health authorities could swell to a flood. The threat of redundancy for the board of an overspending health authority carries much weight when they fear being one of a very few offenders, but little weight when large numbers of other health authorities are also expected to overspend. Indeed, health authority managers may increasingly come to be seen by the populations they serve as failing their local communities *unless* they overspend, given that some health authorities elsewhere will be observed doing just that.

At that point the NHS cash limit would cease to be a ceiling on spending and would become instead a floor: the only certainty being that outturn expenditure would not be beneath that level.

7 SO WHAT?

In summary: the recent 1997/98 NHS White Papers imply that the government is hoping to impose cash limit control over an increased proportion of NHS expenditure while simultaneously removing a major incentive for GPs to co-operate with this attempt. Simply requiring health authorities to keep to their budgeted total costs, including prescribing costs, will on its own be insufficient to guarantee that it is done. It is the GPs who effectively control much of the demand for NHS medicines and other services.

If the abolition of GP fundholding is assumed to be inevitable⁵ and the underlying unpredictability of the demand for medicines is recognised, then the options for avoiding the budget bluff being called, and consequent failure to control public expenditure, boil down to:

- leaving a ‘safety valve’ for some NHS expenditure; or
- creating new incentives to replace those due to be removed in order to get GPs to economise on the claims they make on NHS resources on behalf of their patients.

In the past, the exclusion of FHS medicines expenditure from the NHS cash limit provided a safety valve. Along with emergency hospital care, FHS medicines are a major NHS service to which all members of the public have (near-) immediate access. Given the large element of shared costs in the provision of emergency and elective hospital care, it is impractical to cash limit elective but not emergency care. Either the sum of both must be cash limited or neither of them. The cash limit on hospital expenditure overall, however, has meant that fluctuations and unplanned increases in the demand for emergency hospital care have forced compensating fluctuations and unplanned cuts in elective hospital care, i.e. mainly in elective surgery.

⁵ There are various arguments, unrelated to the question of controlling total NHS expenditure and therefore lying outside the scope of this paper, as to why abolishing GP fundholding may or may not be desirable. The government has focused attention on removing the inequity of a “two-tier health service” and reducing the administrative costs associated with individual GP practices contracting with health care providers. Other arguments concern the effectiveness and efficiency of strategic planning for health care services that may be possible; the need to give a voice to non-medical interests; the skills and training of GPs; and so on. The present paper is, however, focused narrowly on the issue of expenditure control.

Fluctuations and unplanned increases in the FHS medicines bill have not produced such pain hitherto, because the FHS medicines bill has not been cash limited. Instead they may be argued to have had (minor) adverse macroeconomic consequences by producing unplanned variations/growth in public expenditure, and in expenditure on health as opposed to the other areas of government concern. (The adverse macroeconomic consequences are only likely to be material if capital markets interpret failure to control one public expenditure programme, namely health, as a lack of will or ability to control public expenditure more generally). This potential problem with the absence of a cash limit on FHS medicines expenditure hitherto is, however, not mentioned in the 1997/98 NHS White Papers. There the argument appears to be that excluding GPs' prescribing from the overall NHS cash limit may have distorted GPs' treatment decisions for their patients, given that specialist hospital and community health services *are* cash limited. The potential for such distortion does exist, although in practice the extent and consequences of it for patients' welfare are both unknown. The White Papers place an emphasis on 'integrated care'. Placing FHS medicines expenditure in the same cash limited budget as hospital and community health services may make it easier to implement National Service Frameworks and national treatment guidelines that change the pattern of care across traditional primary and secondary care boundaries.

If the government did not consider it important that health care expenditure be constrained to be close to the planned level, it would have no reason to impose cash limits on it. As it chooses to apply cash limits to the health programme, presumably the government intends them to be complied with. On its own terms, therefore, the government needs effective control to be exercised over NHS expenditure. This in turn requires, among other things, that GPs be given incentives to limit the expenditures they initiate. This includes their referrals of patients to other health care services and their prescribing of medicines. As discussed above, evidence from the NHS to date suggests that, as currently specified, the English Primary Care Group approach and its counterparts in Scotland and Wales (and possibly Northern Ireland, depending on future decisions) will not provide adequate incentives to GPs. The strongest incentives for GPs are those which enable at least part of any cost savings achieved by an individual practice that becomes more cost effective in its treatment choices, to be made available to that same practice for allocation to purposes that it determines.

The GP fundholding scheme limited such purposes to those which could be argued to benefit the health care of the GP practice's local population. That incentive appears to have had an effect. Such a practice-based incentive could be introduced with advantage within Primary Care Groups (and their equivalents). This is not a simple matter however, and politically it will be necessary to avoid it appearing to be a return to GP fundholding. Although the English White Paper expects that Primary Care Groups will 'over time' move to setting notional budgets for practices and that practices should be given incentives for staying within them, no details have yet been provided. There are no details about the timetable for such a move, or of how practice budgets might be set, or of how incentives can be made real without threatening equity of access to services for the patients of different GPs or the desire to minimise bureaucracy by avoiding contracts between providers and individual GP practices (a practice with an incentive to economise will want to control its expenditures itself).

There has been much discussion of the real difficulties of determining budgets fairly for individual practices (against which achievement of savings is to be judged) and of allowing for the inevitable variability from one year to the next in the patient needs presenting to a GP. (Smith, 1997, provides a concise overview of the problems). These difficulties suggest that less than 100 per cent of any 'saving' should be made available to the individual practice, to allow (albeit imperfectly) for the fact that some 'savings' will be a random event or the result of an inaccurate initial budget rather than the outcome of deliberate actions by GPs. For example, the practice could retain 50 per cent of the savings, with the other half going to the Primary Care Group or health authority as a whole.

If all of this sounds too problematic, the only alternatives appear to be either to reinstate a non-cash limited 'safety valve' by removing the cash limit from part of NHS expenditure (and FHS medicines have taken this role hitherto) or to run the risk of undermining the, hitherto impressive, control of total NHS expenditure by provoking GPs and NHS managers to call the budget bluff. Such a loss of face (reversing the cash limit extension) or of control would presumably be unthinkable for the government. So, difficult though they may be to design, practice-based incentives are the least undesirable option from the governments' point of view. Voluntary GP fundholding may be 'out', but compulsory GP fundholding may just have to be 'in'.

Appendix

Deviation of outturn expenditure from planned, expressed as a percentage of planned expenditure – NHS England

The sources of the data for both tables are: Department of Health (1996, 1997 and 1998); Department of Health and OPCS (1993, 1994 and 1995); and Department of Health, OPCS and HM Treasury (1991 and 1992).

Table A1 Cash limited expenditure (current plus capital, net of charges and receipts) for hospital, community health, cash limited family health and related services and NHS Trusts, excluding GP fundholders' prescribing costs

	Outturn % deviation from amount planned		
	One year earlier	Two years earlier	Three years earlier
1997/98	+0.92%	+2.58%	+3.30%
1996/97	+0.11%	+0.51%	+0.40%
1995/96	+0.25%	+0.27%	+1.53%
1994/95	-0.06%	+0.74%	-5.48%
1993/94	-0.99%	-5.53%	+4.34%
1992/93	-1.93%	+5.67%	n/a
1991/92	+1.10%	n/a	n/a
Mean deviation	0.77	2.55	3.01

Note: n/a = not applicable

Table A2 Non-cash limited family health services expenditure (net of charges and receipts) plus (cash limited) GP fundholders' prescribing costs

	Outturn % deviation from amount planned		
	One year earlier	Two years earlier	Three years earlier
1997/98	+1.36%	+2.24%	+3.39%
1996/97	-0.04%	+1.52%	+1.51%
1995/96	-0.27%	-0.11%	+3.47%
1994/95	-0.23%	+3.11%	+7.78%
1993/94	+2.68%	+7.46%	+8.89%
1992/93	+8.11%	+9.21%	n/a
1991/92	+5.76%	n/a	n/a
Mean deviation	2.64	3.94	5.01

Note: n/a = not applicable

REFERENCES

Audit Commission (1994). A prescription for improvement. Towards more rational prescribing in general practice. HMSO, London.

Audit Commission (1995). Briefing on GP fundholding. HMSO. July 1995.

Baines D, Tolley K, Whynes D (1997). Prescribing, budgets and fundholding in general practice. Office of Health Economics, London.

Bevan G, Baxter K, Bachmann M (1998). Survey of budgetary and risk management of Total Purchasing Pilot Projects 1996/97. National Evaluation of Total Purchasing Pilot Projects Working Paper, King's Fund, London.

Department of Health (1996). The government's expenditure plans 1996-97 to 1998-99. Department of Health departmental report. Cm 5212. The Stationery Office, London, March 1996.

Department of Health (1997). The government's expenditure plans 1997-98 to 1999-2000. Department of Health departmental report. Cm 5612. The Stationery Office, London, March 1997.

Department of Health (1998). The government's expenditure plans 1998-1999. Departmental report: the health and personal social services programme. Cm 3912. The Stationery Office, London, April 1998.

Department of Health, OPCS (1995). The government's expenditure plans 1995-94 to 1995-96. Department of Health and Office of Population Censuses and Surveys departmental report. Cm 2212. HMSO, London, February 1995.

Department of Health, OPCS (1994). The government's expenditure plans 1994-95 to 1996-97. Department of Health and Office of Population Censuses and Surveys departmental report. Cm 2512. HMSO, London, March 1994.

Department of Health, OPCS (1995). The government's expenditure plans 1995-96 to 1997-98. Department of Health and Office of Population Censuses and Surveys departmental report. Cm 2812. HMSO, London, March 1995.

Department of Health, OPCS, HM Treasury (1991). The government's expenditure plans 1991-92 to 1993-94. Department of Health and Office of Population Censuses and Surveys departmental report. Cm 1513. HMSO, London, February 1991.

Department of Health, OPCS, HM Treasury (1992). The government's expenditure plans 1992-93 to 1994-95. Department of Health and Office of Population Censuses and Surveys departmental report. Cm 1913. HMSO, London, February 1992.

Glennister H, Matsaganis M, Owens P, Hancock S (1994). Implementing GP fundholding: wild card or winning hand? Open University Press, Buckingham.

Health Service Journal (1996, 1997, 1998). The news stories referred to may be found in the following editions of the Journal: 21/11/96, 5/12/96, 9/1/97, 20/2/97, 27/11/97, 4/12/97, 29/1/98 and 12/2/98.

House of Commons (1997). NHS (England) summarised accounts 1995-96. HC 127, Stationery Office, London, 18 July 1997.

House of Commons (1998). NHS (England) summarised accounts 1996-97. HC 923, Stationery Office, London, 21 July 1998.

Likierman A (1986). Squaring the circle: reconciling predictive uncertainty with the control of public expenditure in the UK. Policy and Politics, Vol.14, No.3, July 1986, pp.285-307.

Marks L, Hunter DJ (1998). The development of Primary Care Groups: policy into practice. The NHS Confederation, Birmingham.

Minister for Health and Social Services (Northern Ireland) (1998). Fit for the future: a consultation document on the government's proposals for the future of the Health and Personal Social Services in Northern Ireland. Department of Health and Social Services, Northern Ireland.

Prowle M, Jones T (1997). Health service finance: an introduction. 4th edition. The Association of Chartered Certified Accountants, May 1997.

Secretary of State for Health (1997). The new NHS: modern, dependable. Cm 3807. The Stationery Office, London, December 1997.

Secretary of State for Scotland (1997). Designed to care: renewing the National Health Service in Scotland. Cm 3811. The Stationery Office, Edinburgh, December 1997.

Secretary of State for Wales (1998). NHS Wales: putting patients first. Cm 3841. The Stationery Office, January 1998.

Seitz R, König H-H, Jelastopulu E, Arnold M (1998). Managed care - an option for the German health care system? Office of Health Economics, London.

Smith J, Barnes M, Ham C, Martin G (1998 - in press). Mapping approaches to commissioning: extending the mosaic. King's Fund, London. Reported in: Gold L, Baines D (1998), Integrated care in the UK and Australia, the report of a seminar held at the University of Birmingham Health Services Management Centre on 15 December 1997, and published by the Health Services Management Centre School of Public Policy, University of Birmingham.

Smith PC (1997). Devolved purchasing in health care: a review of the issues. Nuffield Occasional Papers, Health Economics Series, Paper No.2, The Nuffield Trust.

Towse A and Leighton T (1998 - in press). The changing nature of NCE pricing of second and subsequent entrants. In Sussex J and Marchant N (eds.), Risk and return in the pharmaceutical industry. Office of Health Economics, London.

HM Treasury (1998). Modern public services for Britain. Investing in reform. Comprehensive spending review: new public spending plans 1999-2002. Cm 4011, The Stationery Office, July 1998.

Walshe G (1987). Planning public spending in the UK. Macmillan.

Webster C (1996). The health services since the War. Volume II: Government and health care, the British National Health Service 1958-1979. The Stationery Office, London.

Webster C (1998). The National Health Service: a political history. Oxford University Press, Oxford.

West P (1997). Understanding the National Health Service reforms. The creation of incentives? Open University Press, Buckingham.

Recent OHE publications

Our Certain Fate: Rationing in Health Care

by Alan Maynard and Karen Bloor, 1998 (price £5.00).

Competition through Innovation, Innovation through Competition

by Hannah Kettler, 1998 (price £7.50).

Biotechnology and the Economics of Discovery in the Pharmaceutical Industry

by Helen Simpson, 1998 (price £7.50)

Tuberculosis

by Nick Marchant, 1998 (price £7.50)

Economic Aspects of Non-Hodgkin's Lymphoma

by Keith Tolley, Gareth Morgan, Ray Cartwright and Rhys Williams, 1998 (price £7.50)

National Health Service Reorganisation: Learning from History

by Charles Webster, 1998 (price £5.00)

Effective Commissioning: Lessons from Purchasing in American Managed Care

by Donald Light, 1998 (price £7.50)

From Efficacy to Cost-Effectiveness

by Adrian Towse and Mike Drummond, 1998 (price £5.00)

Trends in Economic Evaluation

by Clive Pritchard, 1998 (price £5.00).

Accountable Health Care: Is it Compatible with Social Solidarity?

by Uwe Reinhardt, 1998 (price £7.50).

Managed Care – an Option for the German Health Care System?

by Robert Seitz, Hans-Helmut König, Eleni Jelastopulu and Michael Arnold, 1998 (price £10.00).

Managed Care – a Model for the UK?

by Peter West, 1998 (price £10.00)

Competition and Contestability between Acute Hospitals

by Jon Sussex, 1998 (price £10.00)

