

THE UNIVERSITY *of York*

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

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

Towards Locally Based Resource Allocation in the NHS

Roy Carr-Hill

Nigel Rice

Peter C Smith

DISCUSSION PAPER 159

**TOWARDS LOCALLY BASED RESOURCE ALLOCATION IN THE
NHS**

Roy Carr-Hill, Nigel Rice and Peter Smith

Centre for Health Economics, University of York

Address for correspondence:

Roy Carr-Hill

Centre for Health Economics, University of York, Heslington, York, Y010 5DD.

e-mail: irss23@york.ac.uk.

phone: 01904 432306

August 1998

CENTRE FOR HEALTH ECONOMICS DISCUSSION PAPER SERIES

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

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

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

THE AUTHORS

Roy Carr-Hill is a Reader, Nigel Rice is a Research Fellow and Peter Smith is a Professor of Economics, all located in the Centre for Health Economics, University of York.

ACKNOWLEDGEMENTS

An earlier version of this paper was presented at the Health Economics Study Group meeting in July 1997. The authors gratefully acknowledge the contribution of Steve Birch, the discussant, and other participants. The work is funded by the UK Department of Health, which does not necessarily share the views expressed in the paper.

FURTHER COPIES

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

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

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

ABSTRACT

In a health care system where resources are scarce and the principle of equity is of central concern, mechanisms for the appropriate allocation of resources based on the notion of relating resource use to some concept of need are essential. Two key issues raised in the UK government's White Paper *The New NHS: modern, dependable* are the ability first to define health care budgets at a local level and second to integrate budgets to encompass all relevant aspects of health care delivery. This paper considers the theoretical and practical implications of devolving NHS budgets to primary care groups. The paper advocates the development of a patient-based survey of all NHS health care utilization, which could serve as the basis for integrated global budgets for use at a local level.

Keywords: *resource allocation; health care budgets; primary care; health care need; general practice.*

1. INTRODUCTION

For some years there has been a move in the UK towards a ‘primary care led’ National Health Service (NHS). At the same time, a founding principle of the NHS is that equal access to health care should be available to those in equal need, regardless of individual circumstances. This paper argues a shift of the focus of care towards the primary care level will require increasingly careful definition of “need”. The paper therefore examines the implications of trying to measure need at the level of general practice, or groups of general practices.

The recently elected government has continued the shift in focus towards primary care in the NHS, and the recent White Paper *The New NHS: modern, dependable* signals an increased emphasis on devolution of health care budgets. In this respect, the White Paper builds on existing experience with general practitioner “fundholding”, in which health care budgets for a subset of elective procedures and for prescribing have been devolved to individual general practices which choose to participate. However, the new policy differs in three respects from the fundholding experience: first, the budgetary devolution will be to groups of general practices, in the form of Primary Care Groups, rather than to individual practices; second, the emphasis will be on commissioning and planning services, rather than purchasing; and third, the budget will encompass a much larger proportion of total expenditure than the standard fundholding budget. In this respect, the new arrangements have much in common with the recently evaluated “total purchasing pilots” although even more ambitious in seeking to include primary care expenditure as well as hospital, community and prescribing services.

In this paper we argue that, in order to secure an efficient and equitable distribution of limited health care funds, it is indeed important for devolved budgets to embrace all health care expenditure and not just a subset such as prescribing or fundholding procedures. However, we identify a number of serious practical difficulties that must be overcome if successful devolution along these lines is to take place.

The paper is organised as follows. The next section outlines the importance and role of budgetary management in a health care system where resources are scarce. We then describe the essentials of the present ‘needs’ based formula used to allocate budgets to health authorities. This is followed by a discussion of how health authorities currently secure budgetary control in primary care. In section 3, we consider the move towards devolved budgeting and the implication this will have for resource allocations. Although we endorse the move towards a fully integrated system whereby PCGs will hold a ‘global budget’ from which they will be free to spend on all aspects of health care as they see fit, we acknowledge that – at least in the short term - the problems associated with this will be considerable. We therefore highlight some of the major obstacles that require careful thought and research. In a final section we call for the need for a comprehensive survey of individual NHS utilization designed specifically to address the issues expanded on in this paper and to secure a move towards ‘need’ based primary care led budgetary management.

2. BUDGETS IN THE NHS

Budgets are perhaps the most important formal mechanism for securing managerial control. Emmanuel, Otley and Merchant (1990) set out five roles that budgets usually play:

1. authorization of actions;
2. a means of forecasting and planning;
3. a channel of communication and co-ordination;
4. a means of motivating organizational members;
5. a vehicle for performance evaluation and control.

Whilst these roles may conflict, elements of all five purposes can be discerned in most budgetary systems. In the NHS the principal role of the budget has probably been to signal command over resources, and thereby to secure adherence to public expenditure limits - a mixture of roles one to three.

Expenditure in the NHS can be considered under three broad headings:

- hospital and community health services (HCHS);
- prescribing;
- general medical services, including expenditure on general practices.

This section outlines the systems of budgetary control currently in place in these three categories. See NHS Executive (1997) for further details.

Hospital and Community Health Services

Hospital and Community Health Services (HCHS) expenditure accounted for £20.9 billion in 1996/97, or 76% of all local NHS expenditure in England. These national funds are in the first instance devolved to health authorities with populations of about 500,000. Setting cash-limited annual HCHS budgets for individual health authorities has been central to the objective of restraining government expenditure on the NHS. Of course such budgets could be set on the basis of crude criteria, such as a fixed *per capita* sum, or last year's expenditure plus x%. However, major efforts have been made to make health authority budgets as equitable as possible in terms of responding to variations in need (see below). It is almost certainly the case that the attention to equity has helped to make acceptable the imposition of frequently severe cash limits.

The methodology for setting HCHS budgets for health authorities has been developed and refined over a twenty year period. The general principle is known as weighted capitation because a capitation allocation for each citizen is weighted for a number of relevant factors. Under weighted capitation, a target health authority budget can be expressed as follows:

$$\text{Budget} = \text{PerCap} * \text{Pop} * (1 + a) * (1 + n) * (1 + c)$$

The starting point for the budget is *PerCap*, the national *per capita* sum allocated for HCHS in the Public Expenditure System negotiations. This is first multiplied by the local population size *Pop*. The crude budget this yields is then successively adjusted for the age structure of the population (using the relative age factor *a*), the health care needs of the population over and above age (using the relative needs factor *n*), and the relative costs of local health care (using the relative cost factor *c*). The national averages of *a*, *n* and *c* are zero.

The age adjustment *a* is relatively uncontentious, and reflects the clear propensity of the very young and elderly people to require more health care expenditure than the rest of the population. The needs adjustment *n* has been the subject of intense scrutiny. Current methods involve the use of a set of relative needs indices for different parts of HCHS. These include variables reflecting both health status and more general socio-economic status. For example, the index for acute care, used to distribute 64% of HCHS expenditure, uses the following variables:

- Standardized limiting long standing illness ratio (under 75);
- Standardized mortality ratio (under 75);
- Proportion of economically active who are unemployed;
- Proportion of pensionable age living alone;
- Proportion of dependants in single carer households.

The cost adjustment seeks to adjust for unavoidable differences in costs of capital and labour in different parts of the country, most importantly London and the south east. A separate cost adjustment is calculated for each health authority.

Health authorities are responsible for living within their annual HCHS budgets. In this respect, they face a problem in the sense that HCHS expenditure is heavily influenced by the referral practices of a large number of individual GPs, over whom health authorities have little direct control. Until the advent of fundholding, the major restraining influences have been supply side restraints, in the form of waiting lists for elective procedures and (in extreme cases) a refusal to undertake certain procedures. Nevertheless, even before the 1991 internal market reforms, health authorities were generally successful in keeping within budget limits. This success could not have been achieved without the widespread acceptance by GPs of the need to restrain health care expenditure.

Primary Care: Purchase of HCHS

The introduction of fundholding has allowed health authorities to devolve an average of 15% of their HCHS budget to general practitioners, mainly in the form of standard fundholding budgets. This devolution encompasses a range of common elective procedures and prescribing (see below). However it does not necessarily absolve the health authority of all responsibility for that element of their budget, as the majority of any overspend is likely to have to be met by the health authority. Nevertheless, it was clearly hoped that the devolution would offer GPs a concrete incentive to restrain expenditure for the relevant services, as they were allowed to retain any underspend on their fundholding budget for expenditure on other patient services.

In contrast to the mature budget setting process for health authorities, setting budgets for general practitioners is still in its infancy. An early attempt to develop a needs index for standard fundholding procedures failed (Sheldon et al, 1994). Subsequent guidance from the NHS Executive (1996a) urged health authorities to use some sort of formula to set fundholder budgets, and to ensure that fundholders and non-fundholders were treated equitably. The Executive recommended use of a dampened form of the acute sector index mentioned above (NHS Executive, 1996b), and in practice many health authorities have tended to use similar methods to the NHS Executive weighted capitation approach in setting fundholding budgets. This has entailed weighting each patient on a practice's list by the patient's age, and the health and socio-economic characteristics of the small area in which the patient lives.

The standard fundholding scheme has been augmented by the introduction of "total purchasing", encompassing virtually all HCHS services, including emergencies, and usually involving coalitions of several practices. In contrast to standard fundholding, the total purchasing scheme is the subject of a thorough evaluation (Mays *et al*, 1997). Participating general practices continued to be the formal holders of their own individual budgets, although the architects of the total purchasing scheme envisaged that eventually an entire site would receive a fully integrated budget. Budgeting emerged as a crucial issue in the development of the total purchasing experiment which has yet to be satisfactorily resolved.

Primary Care: Prescribing

In 1996/97 total prescribing costs accounted for £3.8 billion; 14% of all local NHS expenditure. The NHS reforms and the creation of fundholding were intended, in part, to place greater emphasis on cost-effective prescribing in an attempt to control the rise in total prescribing costs. Fundholding budgets include an element for prescribing costs, and fundholding practices have been required to meet prescribing costs out of the fundholding budgets which are allocated to them by their health authority. In contrast, non-fundholders are allocated an indicative prescribing budget. This is a notional budget or target, and penalties for overspending play a much weaker role than for fundholders.

Until relatively recently, budgets for prescribing at the health authority level were largely based on historical costs adjusted for inflation. More recently, allocations to health authorities have moved towards a weighted capitation basis. In 1996/97, for the first time, a very small proportion of the prescribing budgets was based on a needs weighting. After appropriate adjustments for the age, sex and temporary resident characteristics of practices (Roberts and Harris, 1993), a weighting for the proportion of people in the 1991 Census declaring themselves as unable to work due to permanent sickness or disability was applied to calculate health authority allocations (Rice *et. al.*, 1997).

Methodology for devolving health authority prescribing budgets to individual general practices is much less advanced. The indicative prescribing scheme sets implicit prescribing levels for non-fundholding practices, but because of the lack of incentives and penalties to

encourage compliance on the part of the GPs, the scheme has been criticised and has generally failed to control the rise in costs (Walley, Wilson, and Bligh, 1995). Fundholding, although not reducing the total drugs bill, appears to have been successful in reducing the rate of increase of drug costs in participating practices (Department of Health 1994). This appears to be due to the obvious financial benefits to practices of being more conscious of the cost implications of their prescribing patterns.

In the early years of fundholding, allocations to individual general practices were based extensively on an historical cost basis. More recently advice to HAs has been to allocate prescribing budgets based on a capitation system known as known as ASTRO-PU (Age, Sex, Temporary Resident Originated Prescribing Units). These apply weights based on the age, sex and temporary resident structures of practice lists, and are applied to practice populations to derive an overall practice budget. Used in this way, ASTRO-PU represent a crude adjustments for the demands placed on practices due to readily measured characteristics of the practice list (Roberts and Harris, 1993; Lloyd *et al*, 1997).

However, whilst this method of capitation accounts overall for a reasonable proportion of variation observed in prescribing costs across practices (approximately 25%) (Roberts and Harris, p488, 1993), there are wide variations between health authorities and practices. Factors such as out-of-hours services and prescription charge exemptions may account for some of the unexplained variation (Whynes, Baines *et al*. 1996). To incorporate such considerations in the budget setting procedure, initial budgets based on ASTRO-PU are subject to bilateral negotiation between practices and health authorities to ensure that any special needs a practice may encounter are met. Clearly this practice may give rise to incentives for strategic behaviour and inequity in the sense that allocations are not related to a consistent concept of need.

There is great scope for further understanding the mechanisms behind prescription cost variations, particularly at the practice level and in developing suitable models to allocate prescribing budgets on an equitable basis. Whether sufficient variation can be explained by relating costs to measures of need at the practice level is central to the issue of devolving budgets equitably.

Research in the area of prescribing is bedeviled with methodological difficulties which need to be addressed before major inroads can be made into suitable resource allocation mechanisms. Some of the difficulties are common to all sectors of health care, and are addressed in more detail below. Further issues of particular concern in relation to prescribing include:

- Consideration of practice supply characteristics (for example, is prescribing a substitute for other types of health care?)
- The interface between primary and secondary care led prescribing
- A lack of standardization of costs and items prescribed
- Patient prescribing history

- Level of generic prescribing
- Level of repeat prescribing

Primary Care: General Medical Services

HCHS and prescribing account for 90% of NHS expenditure. The major other category of expenditure is general medical services (GMS). GMS embrace most aspects of primary care, including the provision of general practitioners and most of their staff. Only a quarter of the overall GMS budget is subject to national cash limits (GMS “Cash Limited”). The White Paper envisages that only GMS cash limited expenditure will be included in unified budgets in the first instance. This includes funding of practice staff, premises and computers, but not remuneration of general practitioners. Current guidance issued by the NHS Management Executive for setting GMS budgets to the district authorities states that they are “... to be calculated consistent with the methodology for the reimbursement of non-fundholder costs for GMS staff” (NHS Management Executive, 1996). Weighted capitation formulae were used in 1997/98 for allocations for Cash Limited GMS to health authorities, based on age, sex and ICD chapter stratified consultation rates, developed from an analysis of a survey of individual patients from a sample of practices using the MSGP4 survey (Royal College of General Practitioners *et al*, 1995). There is an additional needs adjustment based on the (under 75) standardized illness ratio. At present, we are not aware of any health authority actively seeking to introduce capitation methods for GMS for allocations to individual practices.

Summary

Current resource allocation methods consider a number of discrete sectors of health care. Each sector is considered separately, and different methods have been used to derive capitations in each sector. The favoured methodological approach has been wherever possible to base allocations on empirical data, in particular seeking to use the national average link between “needs” indicators and measures of utilization. As a result, data limitations often determine the methods used. Even though sectoral budgets are derived separately (for example for acute HCHS and psychiatric HCHS), they are often aggregated into a single budget. Usually there is no obligation for budget holders (such as fundholders) to adhere to the individual components of their budget for each sector, so long as they adhere to the total budget.

3. DEVELOPING WEIGHTED CAPITATION BUDGETS IN PRACTICE

In many ways the move towards a primary care-led NHS should make the distinctions between service headings such as prescribing and fundholding procedures irrelevant and unhelpful, as a principal objective of the initiative is to encourage GPs to secure the best health care for their patients subject to budget constraints, regardless of the service heading under which the care is found (UK Government, 1996). Any partitioning into separate budgets runs the risk of artificially constraining 'rational' assessment by the GPs of the alternatives. Furthermore, a global budget offers no opportunity for GPs to shift expenses to services not covered by their budget. For example, the standard fundholding scheme embraced only routine elective surgery, which accounts for just 20% of HCHS expenditure. There has therefore been an incentive for fundholding GPs to refer patients as emergencies, which lie outside the ambit of their budget, thereby transferring financial liability to the health authority.

The existence of such incentives clearly offers the potential for inefficiencies. In principle, therefore, GP treatment decisions can only be made in an undistorted way if (a) budgets capture *all* expenditure caused by GP decisions (b) GPs have complete *freedom* to switch between expenditure headings - that is, separate budgets are not set for specific services. In practice, however, there may be many reasons why this counsel of perfection is neither attainable nor indeed desirable. Amongst the most important reasons for caution are:

- it may be very difficult to set what are perceived to be equitable global budgets;
- at a strategic level it may be perceived that certain functions (say preventative medicine) could be squeezed out by more urgent health care demands, in which case there may be a case for "ring-fencing" the associated budget to ensure that GPs do not neglect the service;
- it may be unreasonable to expect general practices to take on certain health care risks (such as, for example treatment of HIV/AIDS), for which the associated budget might be held at the health authority or even the national government level;
- more generally, the unpredictable variation in local health care needs may be very high, rendering any budgets meaningless;
- GPs may not have the available information or the decision-making skills to make the rational and efficient decisions demanded by the unconstrained budget;
- the management costs associated with devolution may be very high;
- it may be impossible to design satisfactory rewards and sanctions for underspending or overspending budgets.

We nevertheless believe that the devolution of budgets to PCGs is on balance likely to be beneficial, and moreover consider that it is essential that some attempt should be made towards setting budgets for individual general practices. We therefore now consider the practical implications of such a development. The experience of practice-based budgets noted above has highlighted difficulties that were not apparent at the health authority level. These can be summarized as follows:

- the health authority databases of patients registered with GPs usually indicate larger populations than population estimates prepared by the Office of National Statistics - the phenomenon known as “list inflation”;
- there is extremely limited universally available information on individuals – in practice it is confined to age, sex and postcode of residence, leading to the use of small area data as the basis for capitation adjustments;
- with current procedures all patients in a small area are therefore weighted equally, yet there is no reason to suppose that in general patients attending a particular practice are representative of the area in which they live (the so-called “problem of attribution”);
- expenditure variations from expected *per capita* expenditure become very large for small population sizes, such as lists for individual general practices.

In the long run, as patient databases become more reliable and comprehensive, some of these problems may be overcome. However, they present serious obstacles to setting equitable GP budgets for the foreseeable future. We discuss each in turn.

List Inflation

Budgets for health authorities are based on contemporary projections of population produced by the Office of National Statistics. UK population estimates have been the subject of some methodological difficulties, mainly arising from substantial evidence of incompleteness of the 1991 Census of Population, which serves as the basis for current estimates. However, for the purposes of resource allocation to such large areas, the population estimates are generally considered to be adequate.

However, national population estimates are not in general suitable for allocations to smaller areas. Although estimates are produced for local authority wards, these are subject to considerable potential error, particularly when age specific populations are examined. More importantly, if the general practice is the unit of interest, geographical population estimates are inappropriate, as a practice may draw its patients from a variety of wards, and residents of a specific ward may use a variety of general practices.

Therefore the unit of population used to set budgets for individual practices or PCGs must be the practice list. These are maintained by health authorities, and at that level, provide very different population estimates from those suggested by the national estimates. GP lists in England are inflated on average by 5.9%, ranging from 28.9% in Ealing, Hammersmith and Hounslow to -8% in Morecambe.

The general view is that the health authority practice lists are unreliable (because of mobility among young adults and delays in removing list members on death and emigration) and could be dramatically improved (although, of course, there will always remain problems in registering immigrants, temporary visitors, the homeless and refugees). Insofar as reasonably uncontentious budgets can be set for health authorities, this would not be a source of great concern if the inflation were uniform across a health authority. Practice and PCG could simply be scaled down

(or up) to conform to the health authority budget. However, there is evidence of substantial inflation variation within health authorities, with for example list inflation ranging from 3% to 27% amongst individual practices in one London authority. Furthermore, these variations are likely to be even greater within particular age groups, which may exacerbate anomalies in budgets whenever an age weighted capitation formula is used.

Limited Information

The amount of useful information currently universally available about patients on a GP list is minimal, being limited to age, sex and postcode. Its usefulness as a tool for resource allocation purposes is therefore severely constrained. There is no reason why, in principle, GP lists should be subject to such limitations, and the development of the NHS patient number offers the opportunity to ensure that the vast bulk of the population is attached to one and only one practice and that more extensive and reliable information is attached to the record. However, any such improvements will pose formidable practical problems and have significant expenditure consequences. The research agenda is to identify the costs and benefits of proposed enhancements.

If the GP list is to be used as the basis of budget allocations, it will in the future be desirable to incorporate as many routine data items as possible into patient records in order to facilitate the development of a sensitive capitation formula. Any item that is included should in principle be:

- useful - an acknowledged risk factor;
- reliable - not subject to major errors;
- practical - readily collected at reasonable cost;
- universal - available for all patients;
- objective - not subject to substantial variations in judgement;
- up-to-date - capturing contemporary circumstances of the patient;
- free of perverse incentives - in particular, not encouraging GPs to over-treat or under-treat, or to distort reports of health status.

In practice, the range of factors that satisfy these criteria is likely to be quite small. Age and sex clearly satisfy all, and will always form the basis of any system. However, almost all other conceivable items fall down on one or more of the criteria. For example, many measures of health status may be subject to judgement on the part of patient or GP (for example, self reported measures of health such as limiting long term illness (Sutton *et al* 1997)), and may be difficult to keep up to date. Social circumstances, such as living alone, may change from time to time and be difficult to define in any watertight fashion. Economic circumstances, such as employment status are even more difficult to update. One of the more important risk factors may be a measure of past health care utilization. An important research agenda is therefore (a) the extent to which any of these can be used for resource allocation purposes and what effect they would have on allocations, and (b) whether the use of any of them would induce any unintended behavioural responses on the part of GPs or other health workers.

The Problem of Attribution

Hitherto, because of the difficulties of obtaining individual data outlined above, British resource allocation methods have used the characteristics of the small area in which an individual lives - rather than that individual's own circumstances - as the basis of a capitation formula. This has many advantages, most notably the vastly increased volume of data that become available as a basis for resource allocation. However they bring with them the difficulty that the circumstances of the individual may not be typical of the area in which they live, and that the patients from a small area on a particular GP's list may not be representative of that small area as a whole. This is the "problem of attribution".

An associated problem is the adequacy of any small area data used. Data from the Census of Population are available only every 10 years and therefore date rapidly and may suffer from incomplete enumeration. Furthermore, the Office of Population Censuses and Surveys has itself documented the possible bias when using Census data at the smallest area level – the enumeration district, with a typical population of 500 (OPCS, 1993). This arises from the problem of random fluctuations arising from small numbers and is exacerbated by the practice of “Barnardization” (entailing the quasi-random addition of -1, 0, +1 to all counts except basic population). This problem is especially relevant to use of 10% sample tables, and may mean that data from larger areas (such as wards, with average populations of 5,000) may in fact yield more reliable estimates of the aggregate characteristics of individuals living in the small area.

The pitfalls associated with linking small area data to individuals is known as the "ecological fallacy" (Selvin, 1958), under which associations observed at the area level are wrongly inferred to exist at the individual level. The Fourth National Morbidity Survey in General Practice (MSGP4) contains GP consultation data on all patients (approximately 500,000) registered with 60 participating practices over a one year period. Socio-economic data were collected by means of a questionnaire administered to each individual which yielded an 85% response rate. Area of residence postcode linkage allowed Census small area statistics data to be attributed at both the ward and enumeration district level to individual patients.

This permitted an analysis of whether the assumption that enumeration district data reflected the aggregate characteristics of people living in a district better than ward level data (Carr-Hill and Rice, 1995). The mean consultation rates for different categories of social class, tenure status and unemployment status across all individuals were shown to display the expected associations. However, when the enumeration districts and wards in the study were grouped according to the social class distribution, the level of owner occupancy, or the unemployment rate, there was substantial evidence of the ecological fallacy. Moreover, no significant improvement was detected when moving down from ward to enumeration district level.¹

¹ Note, this type of analysis is a *strict* test of the ecological fallacy because precisely the same individuals were compared using first individual and then small area data.

Clearly the extent to which the problem of attribution gives rise to distortions in budget allocations (and consequent incentives to recruit healthy patients from deprived areas) is an important empirical research matter. If the problem of attribution is found in practice to be of relatively minor importance, the need for more comprehensive personal records may be called into question.

Sources of Variations from Budget

Annual health care expenditure on individuals is highly variable and largely unpredictable. Such variation can be considered under five headings:

1. variation that is predicted by the relevant capitation formula;
2. other variation which is in principle predictable (given the individual's characteristics) but which is not captured by the current formula;
3. variation which is due to clinical practice;
4. variation which is due to local health care prices;
5. variation which is random (that is, for practical purposes unpredictable).

In general, only the first of these sources of variation is captured by the budget. The second, predictable source of variation reflects incompleteness or errors in the current formula. There are always likely to be needs factors that cannot be captured in a capitation formula, and yet which are observable by individual general practitioners. If GPs are able to predict that expenditure on a particular patient will vary from the associated capitation payment, the potential for cream-skimming arises (Matsaganis and Glennerster, 1994). Cream-skimming is the process whereby GPs either seek out patients to add to their lists who are expected to have low costs relative to their capitation payment, or seek to deter or remove from their lists patients whose costs are expected to exceed the associated capitation payment. Although the incentive for cream-skimming exists under most systems of health care, there is no evidence that it has hitherto been a major issue in the NHS, even after the advent of fundholding. However, as the budgetary process starts to encompass a larger proportion of health care expenditure, and if sanctions for overspending were to become more severe, then it may assume more importance.

The third source of variation – variations in clinical practice - should properly be ignored by a resource allocation formula, which seeks to capture some “standard” set of clinical practices. The fourth source of variation may need to be incorporated into a budget if there exist unavoidable variations in input prices. In England, a complex “market forces factor” is used to accommodate such variations in hospital and community health services (Institute for Employment Research, 1996).

The fifth, unpredictable source of variation can be considered “random”, in the sense that it defies all attempts at systematic modelling. For an individual patient, unpredictable variations from the annual capitation implied by his or her needs rating are likely to be massive. However, as patients are aggregated into populations, positive and negative variations will start to balance each other, so that the unpredictable *per capita* variation from the capitation budget becomes

smaller. Such aggregation is known as risk pooling.

At the level of the health authority (with a typical population of 500,000) budget risk arising from the random element of variation is unlikely to be a major consideration. However, at the level of the general practice, with typical populations of (say) 6,000, a number of authors have shown that budget risk is likely to be very large (Crump et al, 1991). Martin, Rice and Smith (1997) suggest that, assuming cost per case contracts are used, a typical fundholding practice (population 10,000) has a 1 in 3 chance of incurring expenditure more than 10% away from its acute sector budget, compared to 1 in 400 for a population of 100,000. The extent to which extension of the budget to other services would alter these figures will be an important consideration in the new NHS.

Thus, for whatever reason, the actual expenditure incurred by a general practice is very likely to vary substantially from its budget in any one year. This may lead to:

- low spenders “spending up” to seek to justify their budget;
- high spenders imposing unjustified constraints on treatment and reacting with hostility to the budgeting system;
- patients with identical needs in different practices being treated differently;
- patients with identical needs in the same practices being treated differently depending on the time of year they present (Glazer and Schmueli, 1995);
- budget holding general practices or PCGs negotiating block contracts, thereby transferring the risk to providers;
- budget holding general practices or PCGs taking out insurance with a third party, resulting in an unproductive outflow of funds from the NHS.
-

Some of these responses may be severely dysfunctional.

Given the discussion above, variations from budget might be for some or all of the following reasons:

- the budget formula is faulty or incomplete;
- the referral and treatment policies of the practice differ from the average assumed in the budget formula;
- the practice has negotiated contract prices which differ from the average assumed in the budget formula;
- the pattern of disease amongst practice patients differs unpredictably from the needs-adjusted average assumed in the budget formula.
-

The managerial implications of the four sources are clearly very different.

Clearly improvement in the capitation formula could offer some hope of reducing the problem of health needs variation. Evidence from the US and the Netherlands (Newhouse *et al* 1989, van Vliet *et. al.* 1992) suggests that one possible way of improving the predictive power of the English capitation formulae would be to incorporate data concerning pre-existing clinical

conditions and past health care use of individual patients. Two major difficulties would be associated with such innovations. First, they necessitate the development of objective measures of health status which do not depend significantly on clinical judgement. In practice, this may imply reliance on previous health care expenditure. Second, they may offer a perverse incentive for GPs to increase expenditure on individuals in order to secure a higher capitation fee in the future.

However it is important to recognize that, although the use of a capitation formula is essential, and that some improvement in capitation formulae can be envisaged, no formula - however refined - can capture all the variations in health care utilization. Rather, Martin, Rice and Smith (1997) argue that unpredictable variations in health care needs (which are beyond the control of the GPs) are likely to be the dominant source of variation, suggesting that careful audit of such variations is essential before any managerial action is taken.

In the absence of a perfect prediction of health care needs, there are a number of managerial strategies for handling the inevitable random fluctuations in health care expenditure for small populations (Martin, Rice and Smith 1997). Budgets could be set for periods longer than one year; certain expensive treatments or certain predictably expensive patients could be excluded from the budget; or contracts could be organised to eliminate or share risk. At the same time, if such arrangements totally protect a budget-holder, then the general practitioner effectively no longer has any incentive to restrain expenditure, which negates the original purpose of the budget devolution.

4. TOWARDS A HEALTH CARE USE SURVEY?

Thus far we have considered present allocation mechanisms and some of the practical difficulties encountered using currently available data sources. In this section, we finish by considering how a survey of individual citizens may help to address the methodological problems inherent in current budget development.

It is important to recall that – for practical reasons – budgeting methods in the various sectors of the NHS, such as the acute sector, prescribing and general medical services, are currently developed separately, using different methodologies to derive capitation rates in each sector. Even though there is no need for the budget holder to adhere to each sectoral budget, this discrete approach may in principle give rise to methodological problems in determining fair budgets when, as in the NHS, empirically based capitations are used. The problem arises because there may be significant scope for alternative modes of treatment which imply an element of substitution between sectors – for example, prescribing as a substitute for GP consultations.

Under these circumstances, a clearly better methodology than existing methods would entail basing capitations on the *total* use of NHS resources made by individuals across all sectors, thereby accommodating potential substitutions. The method of setting a global capitation health care budget for an individual patient based on that individual's characteristics might then be as follows. There may be certain reliably measured characteristics of *all* citizens. Let us suppose (purely for illustrative purposes) that they are age, sex, living alone and limiting long-standing illness (ignoring for the moment the difficulties associated with the last two). This might yield a contingency table as follows:

Age	Sex	With long term illness		Without long-term illness	
		Living alone	Not alone	Living alone	Not alone
0-14	Male				
	Female				
15-44	Male				
	Female				
45-64	Male				
	Female				
65+	Male				
	Female				

Each cell would have a tariff, or capitation amount, associated with it (of course some cells might necessarily be empty). Then a general practice's budget would be determined by multiplying the number of patients in each cell by the associated capitation.

Current methods are based on a hybrid approach. Crude individual characteristics – namely age (and in some cases sex) – are used on a contingency basis, as in this example. However the associated capitation is then adjusted up or down depending on the “needs” score derived for the small area in which the patient lives. In principle, to use a pure contingency table approach would offer a more scientifically rigorous (and understandable) method for deriving capitations, which eliminates the problem of substitution, and overcomes the ecological fallacy. However, this example begs the question: how are the capitation weightings to be derived?

The obvious solution in principle is to undertake a sample survey of individual citizens. This would record all possible personal characteristics of individuals in the sample that could be routinely held by health authorities as part of all patients' NHS administrative records. Thus, apart from routine items such as age, sex and address, data such as long term illness and recent use of NHS resources might be included. Considerable attention will have to be given to what social data - such as living alone - are to be recorded, and what mechanisms for maintaining the database are to be used. There will be little point in including data which cannot be updated regularly and reliably. The criteria for inclusion, already noted above, would be: usefulness; reliability; practicality; universality; objectivity; timeliness; and freedom from perverse incentives.

The principal task of the survey would be to record all NHS utilization of the individual over the course of (say) a year. (Note that, as we assume we are interested in NHS care only, we omit private health care expenditure.) The survey might seek to capture inpatient, outpatient and community health care, prescribing and primary care consultations. Such data in themselves would of course be invaluable. However, the main intention would be to cost such utilization in order to identify the tariffs to be entered into the above contingency table.

Statistical analysis could then be undertaken to identify which of the individual characteristics are the most important risk characteristics associated with NHS utilization. Here great care would have to be taken to model any important hierarchies in the data (patients within general practices within health authorities). This is particularly important because of the likely importance of supply effects in determining utilization. That is, the magnitude and pattern of utilization might be heavily influenced by the resources available to and the policies adopted by the relevant health authority and general practice. It is for this reason that (a) great attention should be paid to the chosen sampling methodology and (b) modelling techniques which are sensitive to the hierarchical data structure should be used (Rice and Leyland, 1996).

The most important risk factors would form the basis of the capitation scheme, and the survey results would be used to identify the weight for each risk class. In this respect it may be necessary to use discrete multivariate techniques to “smooth” the contingency table. Moreover, the contingency table approach implicitly seeks to model interactions between risk factors. If

sample size was too small to secure reliable estimates, it may be appropriate to resort to analytic techniques to identify the impact of each factor independently, along the lines of the synthetic estimation methods used by Benzeval and Judge (1994).

It is important to note that, although probably in practice the most satisfactory approach to deriving capitations, the contingency table approach in itself would not eliminate all of the problems noted above. For example, if patient databases continue to be poorly maintained, inevitable distortions will arise. And there would continue to be fluctuation around any average capitations used in the table, offering a continued scope for cream-skimming. These problems must be addressed via other initiatives, such as improved information systems and careful audit to minimize cream-skimming. Also, the method would still be based on empirical use of NHS resources, and would therefore not accommodate need that is currently unmet (or conversely inappropriate use of NHS services). There would also be a need for careful treatment of any impact of local supply or utilization as measured in the survey. There are moreover numerous details that would need to be resolved before any survey could be implemented, such as the sample size, the frequency of the survey, and how utilization would be measured.

Formidable obstacles therefore stand in the way of implementing such a survey, not least its cost. A great deal of preparatory work would need to be undertaken before embarking on such an undertaking, and there is a substantial research agenda associated with resolution of the methodological difficulties noted above. However, as Carr-Hill *et al* (1994) emphasize, the potential benefits such methods would yield are enormous, and we feel that they deserve serious consideration. Not least, they would settle once and for all the debate on whether the richer dataset offered by area level data compensates for the potential distortions to which the use of such data can give rise.

REFERENCES

Audit Commision. *A prescription for improvement: towards more rational prescribing in general practice*. London: Her Majesty's Stationery Office, 1994.

Benzeval, M., Judge, K. The determinants of hospital utilisation: implications for resource allocation in England. *Health Economics* 1994; **3**: 105-116.

Bligh, J., Walley, T. The UK indicative prescribing scheme. *PharmacoEconomics* 1992; **2**: 137-52.

Bogle, S.M., Harris, C.M. Measuring prescribing: the shortcomings of the item. *BMJ* 1994; **308**: 637-40.

Bradlow, J., Coulter, A. Effect of fundholding and indicative prescribing schemes on general practitioners' prescribing costs. *BMJ* 1993; **307**: 1186-9.

Carr-Hill, R.A., Sheldon, T., Smith, P., Martin, S., Peacock, S., Hardman, G. Allocating resources to health authorities: development of method for small area analysis of use of inpatient services. *BMJ* 1994; **309**: 1046-9.

Carr-Hill, R.A., Rice, N., Roland, M. Socio-economic determinants of rates of consultation in general practice based on fourth national morbidity survey of general practices. *BMJ* 1996; **312**: 1008-12.

Carr-Hill, R.A., Sheldon, T. Designing a deprivation payment for general practitioners: the UPA(8) wonderland. *BMJ* 1991; **302**:393-6.

Carr-Hill, R., Rice, N. Is enumeration district level an improvement on ward level analysis in studies of deprivation and health? *Journal of Epidemiology and Community Health*, 1995; 49: S28-S29.

Central Statistics Office Brief (95) 8. London: Her Majesty's Stationery Office, 1995.

Crump, B.J., Cubbon, J.E., Drummond, M.F., Hawkes, R.A., Marchment, M.D. Fundholding in general practice and financial risk. *BMJ* 1991; **302**: 1582-1584.

Davey Smith, G. Second thoughts on the Jarman index. *BMJ* 1991; **302**: 359-60.

Department of Health. Evidence to the House of Commons Health Committee. *Priority setting in the NHS: inquiry into the NHS drugs budget*. London: Her Majesty's Stationery Office, 1994.

Edwards, C., Metcalf, D., Burn, A., Watson, K., Seward, F.C.N., Jephson, M.H. *et al.* Influence of patient age on drug costs; an investigation to validate the prescribing unit. *International Journal of Pharmacy Practice* 1991; **1**: 73-8.

Emmanuel, C. R., Otley, D.T., Merchant, K. *Accounting for management control*, London:

Chapman and Hall, 1990.

Forster, D.P., Frost, C.E.B. Use of regression analysis to explain the variation in prescribing rates and costs between family practitioner committees. *British Journal of General Practice*, 1991; **41**: 67-71.

Glazer, J., Shmueli, A. The physician's behaviour and equity under a fundholding contract. *European Economic Review* 1995; 39: 781-785.

Harris, C.M., Roberts, D., Lloyd, D.C.E.F. Practice list inflation [Letter]. *British Journal of General Practice* 1996; January: 50.

Healey, A.T., Yule, B.F., Reid, J.P. Variation in general practice prescribing costs and implications for budget setting. *Health Economics* 1994; **3**: 47-56.

Institute for Employment Research, *Labour market forces and NHS provider costs: final report*, Warwick: Institute for Employment Research, 1996.

Jarman, B. Identification of under-privileged areas. *BMJ* 1983; **286**: 705-9.

Lloyd, D.C.E.F., Harris, C.M., Clucas, D.W. Low income scheme index: a new deprivation scale based on prescribing in general practice. *BMJ* 1995; **310**: 165-70.

Lloyd, D.C.E.F., Harris, C.M., Roberts, D. Unemployment and prescribing [Letter]. *British Journal of General Practice* 1994; **44**: 331-2.

Lloyd, D.C.E.F., Scrivener, G. Growth and relative stability of the proportion permanently sick in English family health service authorities, 1981-91. *Journal of Epidemiology and Community Health* 1995; **49**: 222-223.

Lloyd, D.C.E.F., Roberts, D.J., Sleator, D. Revision of the weights for the age sex temporary resident originated prescribing unit. *British Journal of Medical Economics*, 1997; **11**: 81-85.

Martin, S., Rice, N., Smith P., Risk and the GP budgetholder. *Discussion Paper 153*, Centre for Health Economics, University of York, 1997.

Matsaganis M., Glennerster, H. The threat of 'cream-skimming' in the post-reform NHS. *Journal of Health Economics* 1994; **13**: 31-60.

Mays, N., Goodwin, N., Bevan, G., Wyke, S. *Total purchasing: a profile of national pilot sites*, London: Kings Fund, 1997.

Morton-Jones, T., Pringle, M.A.L. Prescribing costs in dispensing practices. *BMJ* 1993a; **306**: 1244-6.

Morton-Jones, T., Pringle, M. Explaining variations in prescribing costs across England. *BMJ*

1993b; **306**: 1731-34.

Newhouse, J.R., Manning, W.G., Keeler, E.B., Sloss, E.M. Adjusting capitation rates using objective health measures and prior utilisation. *Health Care Financing Review* 1989; **3**: 41-54.

NHS Management Executive EL (94) 72. *Purchasing and prescribing*. Leeds: Department of Health, 1994.

NHS Management Executive, *HCHS Revenue Resource Allocation to Health Authorities. Weighted Capitation Formulas*; Leeds: Department of Health 1997.

NHS Executive. *General practitioner fundholder budget-setting: the national framework*, NHS Executive Letter EL(96)55, Leeds, 1996(a).

NHS Executive. *Financial management and GP fundholding*, NHS Executive Letter EL(96)82, Leeds, 1996(b).

NHS Executive. *Progress report of the Resource Allocation Group, 1995 and 1996*. NHS Executive Catalogue no. 96FP0041, Leeds 1997.

Prescribing Research Unit. *The allocation of prescribing budgets by Regional Health Authorities to their Family Health Service Authorities*. Proposal to the Department of Health. Prescribing Research Unit, November 1993.

Prescribing Research Unit. *The allocation of prescribing budgets by Regional Health Authorities to their Family Health Service Authorities*. Supplementary Report to the Department of Health. Prescribing Research Unit, November 1993.

Pringle, M., Morton-Jones, A. Using unemployment rates to predict prescribing trends in England. *British Journal of General Practice* 1994; **44**: 53-6.

Rice, N., Carr-Hill, R.A. *Predicted power of models derived from MSGP4*. Report to NHS Management Executive March 1996.

Rice, N., Carr-Hill, R.A., Roberts, D., Lloyd, D. Informing prescribing allocations at district level in England. *Journal of Health Services Research and Policy* 1997; **2**: 154-159.

Rice, N., Carr-Hill, R.A., Jenkins-Clarke, S. Consultation workload. Draft report to NHS Management Executive, March 1997.

Rice, N., Leyland, A. Multilevel models: applications to health data. *Journal of Health Services Research and Policy* 1996; **1**: 154-164.

Roberts, S., Harris, C.M. Age, sex and temporary resident originating prescribing units (ASTRO-PU): new weightings for analysing prescribing of general practice in England. *BMJ* 1993; **307**: 485-88.

Royal College of General Practitioners, Office of Population Censuses and Surveys, Department

of Health *Morbidity statistics from general practice: fourth national study 1991-92*, London: HMSO, 1995.

Scrivener, G., Lloyd, D.C.E.F. Two proxies for morbidity in the 1991 UK Census: permanent sickness and limiting long term illness. *Journal of Epidemiology and Community Health* 1996; **50**: 223.

Selvin, H. Durkheim's suicide and problems of empirical research. *American Journal of Sociology* 1958; **63**: 601-19.

Sheldon, T.A., Smith, P.C., Borowitz, M., Martin, S., Carr-Hill, R.A. Attempt at deriving a formula for setting general practitioner fundholding budgets. *BMJ* 1994; **309**: 1059-64.

Suckling, J., Carr-Hill, R.A. *The interface between Primary and Secondary Care*, Report to the NHSE, University of York, Centre for Health Economics, 1994.

Sutton, M.A., Carr-Hill, R.A., Gravelle, H., Rice, N. *Potential biases in estimating the determinants of medical-care utilisation using self-report measures: an investigation of limiting long-standing illness and General Practitioner consultations*. Presented to: Health Economics Study Group (HESG) 1997; July: York.

Talbot, R.J. Underprivileged areas and health care planning: implications of use of Jarman indicators of urban deprivation. *BMJ* 1991; **302**: 383-6.

UK Government, *Choice and opportunity. Primary care: the future*, London: HMSO, 1996.

UK Government, *The new NHS: modern, dependable*, London: HMSO, 1997.

van Vliet R.C.J.A., van de Ven W.P.M.M. Towards a capitation formula for competing health insurers. An empirical analysis. *Social Science and Medicine* 1992; 34(9): 1025-48.

Walley, T., Wilson, R., Bligh, J. Current Prescribing in Primary Care in the UK: Effects of the Indicative Prescribing Scheme and GP Fundholding. *PharmacoEconomics* 1995; **7**: 320-31.

Whynes, D.K., Baines, D.L., Tolley, K.H. Explaining variations in general practice prescribing costs per ASTRO-PU (age, sex, and temporary resident originated prescribing unit). *BMJ* 1996; **312**: 488-9.