King's Fund

Resource Allocation within Health Authorities

Lessons from total purchasing pilots

Gwyn Bevan
Reader in Public Policy Analysis
Department of Operational Research
London School of Economics and Political Science

ish Square
London W1M 0AN

'Total purchasing' is potentially the most significant development in NHS purchasing of health services since the introduction of general practitioner fundholding in 1991. It offers fundholding practices the opportunity to purchase all of the health care not included in fundholding for patients on their lists.

The study of 53 TPPs, which went 'live' in April 1996, found that setting budgets was one of the most serious problems they faced in their preparatory period. This report reviews these problems and suggests how they might be resolved through new kinds of analysis. This offers lessons for developing locality purchasing in the NHS.

Published by
King's Fund Publishing
11–13 Cavendish Square
London W1M 0AN

ISBN 1 85717 176 4

Distributed by Grantham Book Services Limited
Isaac Newton Way, Alma Park Industrial Estate
GRANTHAM, Lincolnshire NG31 9SD
Tel: 01476 541 080 Fax: 01476 541 061

The Total Purchasing National Evaluation Team (TP-NET)

The national evaluation of total purchasing pilots in England and Scotland is a collective effort by a large consortium of health services researchers. The study is led by the King's Fund Policy Institute, but also involves the National Primary Care R&D Centre; Universities of Edinburgh, Bristol, Southampton, York and Birmingham; the London School of Hygiene & Tropical Medicine; and the London School of Economics and Political Science. More information about the evaluation is available from: Nick Goodwin, King's Fund Policy Institute, 11–13 Cavendish Square, London W1M 0AN.

Acknowledgements

This paper is based on a survey of lead GPs by Sally Wyke and Jacinta Lee at the University of Edinburgh; and a survey of health authorities by Gwyn Bevan. We are very grateful to those GPs and staff in health authorities who took time to reply to our questions.

An earlier draft of this was circulated to the 300 from England, Scotland and Wales who came to workshops on budget setting for Total Purchasing Pilots in November/ December 1996. These workshops were organised by Regional Offices of the NHS Executive. I am grateful to Doreen Hounslea (North West), Christina Craig (South and West), Peter Lemmey (North Thames) and Paul Gardner (West Midlands) and their staff for organising these workshops.

I am grateful to Max Bachmann, Jennifer Dixon, Howard Freeman, Karen Hancock, Brian Maynard-Potts, Nick Mays, Colin Sanderson, and Trevor Sheldon, for comments on an earlier draft. The usual disclaimer applies.

This work draws on collaborative research over twenty years into resource allocation in the NHS. I am indebted to colleagues on the Resource Allocation Working Group of the Welsh Office (1987–90), the Medical Practices Committee of England and Wales (1993–94), and the South and West Sub-Health Authority Working Group on Technical Target Setting Issues (1997). I am particularly grateful to the chairs: John Wyn Owen, John Ball and Mark Callingham. I am grateful to many others in universities and the NHS who are too numerous to mention individually. I would, however, like to mention my gratitude to John Perrin, Nick Mays, Walter Holland and Trevor Sheldon. It is with great sadness that I record debts I owe to Brian Abel-Smith and Dani Bevan who have died.

The national evaluation was commissioned and funded by the Department of Health. However, the views expressed in this paper do not necessarily represent the policy of the Department.

The Author

Gwyn Bevan is a member of TP-NET, Reader in Public Policy Analysis at the London School of Economics & Political Science, and a member of the NHS Executive's Technical Advisory Group on Resource Allocation.

Contents

Boxes, Tables and Figures

Summary

Total Purchasing Pilots (TPPs) are general practices in which GPs have opted to extend fundholding and have scope to contract for virtually all hospital and community health services. TPPs are pilots and have no statutory basis. The study of 53 TPPs, which went 'live' in April 1996, found that setting budgets for the financial year 1996–97 was one of the most serious problems they faced in their preparatory period (1995–96). All authorities responsible for TPPs faced problems with basic data on activity and expenditure. Most authorities have sought to set budgets for TPPs using a capitation formula. This paper describes the problems encountered in setting budgets for TPPs, and suggests how they might be tackled. It argues that this is relevant to developing locality commissioning in the NHS, and that this development offers scope for improving equity and efficiency in use of resources within health authorities.

The first objective in applying capitation formulae is to discover whether the distribution of resources to populations is equitable and efficient. This entails resolving a number of definitional problems. Which risks ought formulae to aim to capture in weighting populations? Which cost variations ought to be included in the formula? The paper discusses these questions and how they relate to the second objective, which is to make the future distribution of resources more equitable and efficient.

One model for setting budgets for TPPs is offered by the national processes in which Government Health Departments allocate resources to health authorities. This process applies standard methods consistently at authority level. This largely enables changes in allocations to be made without entering into negotiations. This paper identifies a number of problems in seeking to replicate this process for TPPs. TPPs' populations are smaller and built up from general practices; and TPPs (unlike health authorities) have opted to purchase only a subset of hospital and community health services. Problems have arisen when formulae indicate that allocations ought to be changed, which, in effect, means transferring resources for purchasing between TPP and its authority. This paper offers an approach that can contribute to, and illuminate, these negotiations.

The approach suggested here is to follow common practice and estimate 'target' allocations for a TPP based on a capitation formula, and the TPP's estimated past spend. This is likely to give reliable estimates of target spend provided populations are greater than 50 000. These target estimates are likely to differ significantly from estimates of past spend. This paper proposes that the reason for such differences be investigated. The first question is whether the cause is differences in volume or price. Differences in volume could be caused by increased risk in the population or variations in medical practice. This analysis can also be done for the practices which make up the TPP and for TPPs with populations below 50 000. At the level of single practices (or populations of 10 000 or below) clinical information on hospital admissions (on types of conditions and procedures) can be used to assess whether, for example, high admission rates are likely to reflect increased risk or medical practice styles.

If, for example, a TPP has higher volumes of admissions than are expected (based on the capitation formula), in which categories are these high volumes concentrated? Are they in categories of admission of high clinical variation and where inappropriateness is known to be high? If so, this suggests that they may be results of variations in medical practice and volumes ought to be reduced. But, if on the contrary, they are in categories of admission known to exhibit low variation and have high levels of appropriateness, then this suggests the formula is not capturing adequately the high risk of the TPP's population, and the target is too low.

1. Introduction

Nature of Total Purchasing Pilots

The UK National Health Service (NHS) is mainly funded by taxation and offers universal coverage to the population. The NHS reforms introduced in 1991 separated purchasers from providers of secondary care (hospital and community health services)[1]. Purchasers are in effect local insurers that are funded for defined populations and contract with providers. There are two kinds of local insurers: health authorities and General Practitioners (GPs) who have opted to become GP fundholders (GPFHs). Populations of authorities (in England and Wales[2]) are defined by administrative areas, and those of GPFHs by patients who have chosen to register with GPs for primary care.

GP fundholding began on a limited scale, with a small proportion of the population in GPFHs, and a restricted range of services[3]. GP fundholding has since grown considerably in the scale of population covered, and there has also been some extension in the range of services included[4]. The most dramatic extension is where GPFHs have opted to become Total Purchasing Pilots (TPPs) where, in principle, they may take over responsibility as an insurer for virtually all hospital and community health services. TPPs are pilot organisations without a statutory basis (unlike GPFHs and authorities)[5]. TPPs are able to choose which services they will add to those for which they are responsible as GPFHs, and are able to 'block back' other services to their authority. Authorities and TPPs developed their own methods of setting TPP budgets. Teams of researchers are evaluating TPPs. This paper is one outcome of the study of Total Purchasing Pilots (TPPs) by the National Evaluation Team (TP-NET).

Problems with setting budgets

It was originally intended that GP fundholder budgets would be derived using a capitation formula[6]. This would be consistent with the way authorities' budgets are set: with reference to past spend and a target spend derived by a formula based on capitation. Because of methodological problems, however, no capitation formula was used initially for GP fundholder budgets. These were based solely on estimates of past use of services at current prices and revealed considerable variations in spend per capita. This suggested inequities between fundholding practices. It is common practice to use past spend as the start point for setting budgets. Problems were caused for later years by the lack of capitation targets[7] to indicate fair shares between different GPFHs, and between GPFHs and non-fundholding practices:

- There were accusations that GP fundholding introduced a 'two-tier' NHS, and undermined equity[8].
- There were technical problems over taking account of changes in inflation, growth money, increases in emergency admissions (which lay outside fundholding) and any other adjustments to authority allocations (e.g. extra allocations for waiting lists). In principle, these technical problems can be addressed by using past spend as an indication of GPFHs' share of the authority's total allocation.
- There was a problem over setting future budgets to generate appropriate incentives. If GPs made changes that resulted in savings, then this raised the question of how much of those savings they should be allowed to keep in next year's budget[9].

Deriving a capitation-based target based on the insured population helps set future budgets in ways which promote:

- equity by applying the same principles across GPFHs and authorities,
- and efficiency by providing a benchmark for future allocations independent of past use or subsequent changes in allocations to GPFHs.

Because of the many difficulties created by setting budgets for GP fundholders based on their estimated past use of services, most authorities sought to develop capitation target allocations for their TPPs. That in turn, however, also led to many problems. These are summarised in Box 1.

Budgets for GPFHs and TPPs are in effect determined through subtracting their allocations from the host authority's allocation. The study of 53 TPPs, which

went 'live' in 1996–97, found that setting budgets for that financial year (which began in April 1996) was the most serious problem that they faced. Two sites withdrew because of problems in setting budgets. Lead GPs were sent a questionnaire in June 1996 (with reminders in August 1996). They reported that 14 out of 40 had not agreed a budget. The survey of health authorities sent out in August 1996 found that four out of 30 authorities said that they had not yet had an agreed budget. Twelve authorities said that they believed that delays in budget setting were likely to have limited what their TPP was aiming to achieve in 1996–97.

Box 1 highlights problems, which caused delays. Two of these are due to the pilot nature of TPP. Unlike GPFH there is no deadline by which budget offers for TPPs have to be made[10]. Nor is there a formal process for resolving disputes between the TPP and the authority over what the budget ought to be[11]. This paper considers below problems, which would persist, if TPPs lost their pilot status. These are problems of measuring variation in equity and efficiency in use of resources within authorities, and trying to create greater equity and increased efficiency. Creating localities within authorities will reveal such variations and lead to pressure to do something about them. This is why this paper argues that, despite problems in setting budgets for localities, this offers new potential for developing equity and efficiency in the NHS. This potential comes not from provider competition in an internal market, but collaborative investigation of reasons for variations in provider costs and in purchaser volumes. This approach may also be relevant to other countries[12].

Box 1: Reasons for delays in setting TPP budgets

- Problems with obtaining accurate data on past use of health services by practice populations
- Variation between TPPs in definition of services to be included in TPP
- No deadline for setting TPP budgets.
- No process for resolving disputes over budgets between the TPP and the Health Authority
- Formula gives different allocation from estimated past expenditure

Source: TP-NET surveys of health authorities and lead GPs in 1996.

Importance and relevance of total purchasing pilots

There are three main ways of paying providers for health care. These are by basing future budgets on:

- past expenditures, or
- future volumes, or
- population covered.

The first lacks any long-term justification. The US has tried volume based funding through insurance, but is now increasingly seeking alternatives through 'managed care', of which the family of Health Maintenance Organisations (HMOs) is particularly important[13]. HMOs are funded by capitation; and one leading organisational form of HMO is that in which primary care physicians act as gatekeepers to specialists. TPPs and GPFHs are UK variants of the US HMO. Authorities may be regarded as HMOs without gatekeepers: GPs are in principle given clinical autonomy to refer patients to any hospital; and neither hospital doctors nor GPs are employees of the authority. This looks to be a structural weakness[14] and offers an explanation of the recurrent problem most authorities face towards the end of each financial year. Trusts 'overperform' by reaching volumes agreed in contracts with authorities before the end of the year, but authorities have no extra money to pay costs of extra cases[15].

This background suggests that given a purchaser/provider split for secondary care, there is logic to integrating purchasers' responsibilities for insurance with decisions on referrals. This in turn suggests organising purchasing in some way around GPs[16]. There are three obvious options:

- for authorities to involve GPs in locality purchasing, or
- for GPs to become insurers (as in TPPs), or
- for both models to develop alongside each other in the same authority.

Whichever option is chosen, it will be necessary to set budgets for localities within authorities and experience of, and lessons from, setting budgets for TPPs will be relevant. If US experience is a guide, this will also be relevant to insurers (including governments) in other countries in developing ways of organising health services within a system of finance by capitation.

Once information is known on variations in resources per capita in the NHS, this creates pressure for more equitable allocations. This is to be expected, as the rationale for creating the NHS was that access to health care would no longer depend on where people happen to live[17]. Thus right from the start of the NHS, as GPs were paid on the basis of their list size, there was a policy to distribute GPs equitably[18]. Little was done about equitable allocations to hospital and community health services until health authorities were created and spend on services could be related to populations[19]. A year after these authorities had been created, the Department of Health commissioned the Resource Allocation Working Party (RAWP) to address inequity between them[20]; and other UK countries commissioned their own working parties with similar terms of reference. Each country has since distributed resources to health authorities with reference to capitation formulae[21]. There has been evidence of territorial injustice *within* authorities from various studies of variations in hospital admissions by small areas[22] and GP rates of referral[23]. Nothing, however, was done to address the intra-authority problem until the developments of GP fundholding and TPP. Once inequity is revealed in these ways, there is pressure on the Government and authorities to remedy them. Developing locality-based purchasing will put this issue on the agenda and will offer a means to address inequalities within health authorities.

Structure of paper

The rest of this paper consists of three main chapters on objectives, methodological problems and policy choices, and an outline of a new approach.

The chapter on objectives considers the two purposes of using a capitation formula:

- to discover whether the distribution of resources appears to be equitable and efficient, and, if not,
- to make appropriate changes.

This chapter emphasises that the purpose of using capitation formulae is to indicate future changes in allocations over time. It is not to give perfect immediate answers to what allocations ought to be for the next year. Where, for example, there appear to be large differences in targets and past spend within authorities, approximate methods will be adequate in indicating redistributive changes over the next few years. It argues that it is advantageous to use a different approach in allocating resources *within* an authority to that used nationally in setting allocations *between* authorities.

The chapter on methodological problems uses results from surveys of authorities and lead GPs. This gives evidence of approaches taken to, and experiences and difficulties of, setting TPP budgets for 1996–97. The chapter relates this evidence to intrinsic methodological problems in resource allocation and indicates policy choices that need to be made.

Capitation formulae work by multiplying three sets of statistics:

- data on *populations*: numbers, age and other measures of risk;
- *risk rating*: how age and other measures of risk produce changes in use of volumes of services; and
- *costs or prices*: how to translate data on volumes of different types of care into resources.

All three cause problems, but there is a tendency to focus on problems of risk rating, and pay less attention to difficulties in deriving good data on populations and problems caused by variations in costs. The methodological chapter considers problems with all three sets of statistics.

The final chapter suggests that whilst risk rating to TPPs of populations can give reliable indicators of target spend for populations of 50 000 (and above), a new approach can be used to investigate reasons for differences between estimates of past and target spend for smaller populations. This approach focuses on examining the reasons for differences between these two estimates. It can be used for both the practices that make up the larger TPPs and the smaller TPPs. If a TPP, for example, is estimated to have higher spend in the past than its future target, it is important to know whether this is caused by differences in *volume* or *price*.

If the cause is differences in *price*, it is important to understand whether these are:

- unavoidable by the provider (e.g. in labour costs), or
- caused by variations in efficiency, and hence whether the TPP should have an allocation adjusted for differences in prices.

If the cause is differences in *volume* that means that rates of treatment are different those assumed by the capitation formula to adjust for relative risk. For small populations there are limits to what can be achieved by refining measures of risk. The new approach advocated here is to use clinical information on hospital admissions: on diagnoses and procedures. This information can indicate whether, for example, increased levels of admissions are attributable to:

- the formula underestimating the risk of the population, or
- patients and doctors responding differently to risks which are adequately captured by the formula.

2. Objectives

Distributing a cash limit to improve equity & efficiency

The Resource Allocation Working Party (RAWP)[24] was required by its terms of reference to establish a method of distributing resources which was 'responsive objectively, equitably and efficiently to relative need'. The Working Party interpreted the underlying objective as being 'to secure through resource allocation that there would eventually be equal opportunity of access to health care for people at equal risk'. What runs through the original terms of reference and this interpretation is the aim of developing a formula based on objective data to distribute fairly a fixed sum on a per capita basis to authorities weighted for relative need or risk. It is clear that this is *not* a system in which 'money can follow the patient' (i.e. there is no extra money for treating more cases). The aim is to finance relative claims on a fixed sum, and not total costs from demand as determined by doctors. There are, however, two shifts in emphasis from the terms of reference to the interpretation of their meaning, which merit amplification. These are the use of 'need' or 'risk', and the search for equity and efficiency. They are relevant to the problems of setting budgets for TPPs and localities where these contract in some way with providers.

Risk, need and social deprivation

Using 'need' as the basis for allocating resources suggests that those resources be allocated to the 'needy'. The poor in inner cities appear to be needier than the rich in retirement areas. This may be why it is often argued that resource allocation formulae ought to account of 'social deprivation' as well as age (and sex). This begins to suggest that a formula driven by 'needs' ought to compensate for 'social deprivation'. There are, however, problems with this argument[25]. Spending money on health care does not directly help the poor. (The best way of doing this is to increase their income.) The consequences of poverty as manifested in ill health may also be better remedied by attacking the cause (e.g. supplying or improving housing). There is also a tendency to assume a rural Arcadia[26] and that deprivation is exclusively an inner-city problem[27]. To try to avoid confusion generated by vague appeals to account for 'social deprivation,' this paper uses the term 'risk' rather than 'need'. In *principle*, the fundamental question is:

- Which indicators best measure variations in risks of requiring health services by age and sex?

Since the RAWP Report, all national formulae have included the Standardised Mortality Ratio (SMR), as a proxy for additional risks not captured by age (and sex). Hence the fundamental question is the extent to which this combination adequately captures variations in risk. The point here is not whether mortality (as measured by SMRs) reflects all risks (such as those of chronic diseases and elective admissions), but whether these risks vary in any systematic way additional to variations captured by age and sex [28].

As the RAWP report observed, the SMR is correlated with all other indicators of deprivation[29]. Indeed, the widening over time of the social class gradient of mortality suggests that the SMR has become a more sensitive indicator of deprivation now, than it was when it was first proposed twenty years ago[30]. Hence, in *practice*, over the last twenty years in the UK, the two main questions have been:

- What weight ought to be given to the SMR in resource allocation formulae: i.e. if an authority has a SMR 10 per cent higher than the national average, should it, as the RAWP report assumed, receive 10 per cent more money?

- Second, to what extent are problems caused by the poverty associated with indicators of social deprivation a justification for extra resources for health care beyond those captured by age, sex and SMR?

There are two further issues which need to be considered: how to account for age, and does sex matter?

National methods have always accounted for age by using national data on average utilisation rates by

age group. This produces, as expected, high use at birth and for the over-75s, and in particular for the over-85s. Although this has not been controversial (as compared with the debate over SMRs). There are still questions over whether actual use by age group reflects risks by age which would benefit from health services: for example, is there in practice ageism in decisions on treatment and should the young have priority in use of scarce resources for health care[31]?

It is clear that there are different utilisation rates by sex (with women having higher use than men during their child bearing age). If, as we would expect, the ratio of men to women is the same by age group in different areas, then accounting for age alone will be adequate in capitation formulae. Work in the South and West has shown that taking account of sex makes a difference of the order of 2 per cent[32].

Small area variations in utilisation: risk or medical practice?

It is well known that there are large variations in utilisation of services per capita across small areas (of 10 000 population) after having taken account of differences in age and sex. Two quite different lines of investigation have been pursued to offer an explanation of these variations[33]. Having standardised expected rates for differences in age and sex by small area, it has been hypothesised that these variations are caused:-

- *either* by differences in supply and in risk (e.g. as measured by SMR and census variables);
- *or* by differences in medical practice through examination of specific surgical procedures and medical conditions.

The York Study[34], for example, examined small area variations in utilisation (having standardised by age and sex), sought to control for different access to supply, and to explain residual variations through risk factors. In this way this study appears to offer answers to the two main questions in resource allocation formulae (about the weight to be given to the SMR, and what other risk factors ought to be included). Although this has generated credible sets of risk factors, these models only explained about half the total variation. Analysis of small area variations in GP fundholding procedures produced findings that did not make sense in terms of risk (see below).

Studies of small area variations in utilisation (having standardised by age and sex) of specific conditions and procedures show that for some types of admission (e.g. hip fractures) there is little small area variation, but for others (e.g. tonsillectomy) there is large variation. These studies attribute most variation not to differences in risk of populations, but in demands by patients on doctors, and doctors' responses to these demands. For tonsillectomy, for example, one GP may refer most patients with tonsillitis, and another a few; and, one surgeon may operate on most of those referred, and another on a few[35].

There are four implications of the research into medical practice for resource allocation:

- It is not the purpose of resource allocation formulae to explain *all* variations in use of resources. This would merely fund the status quo.
- Formulae should be directed at those risks for which there is capacity to benefit and not those risks for which there is great uncertainty over whether patients will benefit[36].
- A capitation formula is based on expected rates of treatment (for age and other risk factors). Actual volumes of treatment for specific conditions and procedures will be different from expected volumes (from multiplying the formula's expected rates by populations).
- Analysing variances between actual and expected volumes by type of admissions can indicate whether these are likely to be due to the formula inadequately measuring risk, or variations in medical practice.

Variations in costs and volumes

The Resource Allocation Working Party was aware that its terms of reference, by including equity and efficiency, put them on the horns of a dilemma. If an authority is using a hospital which has higher costs than the average, then using a funding formula based on average costs would mean that this authority would lack the purchasing power to pay for the equitable levels of health care assumed by the formula.

But, the remedy of increasing the authority's allocation would generate perverse incentives by rewarding inefficiency.

It is not possible to achieve two objectives (equity and efficiency) with only one policy instrument (funding by capitation)[37]. The final chapter of this paper advocates analysing the cause of differences in estimated past actual spend as compared with a future target spend into differences in *price* or *volume:*

- Correcting differences in *price* may require action by the provider.
- Differences in *volume* may be due to variations in underlying *risk*, and require more resources, or be
- due to *medical practice variation*, and require action by GPs and hospital doctors.

Redistribution

The implications of differences between estimates of future target and past actual allocations for the TPP (and the rest of its authority) are that either the TPP gains at the expense of the rest of its authority or vice-versa. In the survey of authorities with TPPs, nineteen out of thirty reported delays in setting budgets. Ten of these attributed this to the fact that the capitation formula produced an estimated target spend which was so different from the estimated past actual spend, that basing an allocation on the target was unacceptable[38]. How can this problem of redistribution be presented and managed over time to ease its resolution?

One approach used by some authorities (and recommended by the Department of Health in Scotland) is to reframe the distributional problem so as to avoid casting this as the authority versus the TPP. Suppose, for example, applying a capitation formula shows the TPP's estimated past expenditure to be 80 per cent of its estimated future target. This may lead into a struggle for resources between the TPP and its authority. It may be more fruitful to divide the authority into a number of localities, with one (or more) of them as TPP(s). This will then give an indication of variations in estimated expenditures around targets across localities within the authority. Figure 1 indicates a possible outcome from such an analysis: with locality A currently receiving only 60 per cent of its target allocation, the TPP receiving 80 per cent, and localities B, C and D receiving 90 per cent, 110 per cent and 120 per cent. This redefines the task of redistribution in three ways.

- It is no longer setting TPP against authority, but raising for the whole authority the question of redistribution between localities.
- The focus will be on addressing the extreme variations. Thus, if the outcome were as shown in Figure 1, the initial focus would be on moving resources from locality D to locality A, and not to change the TPP's allocation.
- Starting with the extremes and managing changes over a number of years, also gives time to improve understanding of the reliability of the capitation formula[39]. This will enable those involved to become more confident about what further changes might be made once the extremes have been reduced. This may show whether anything ought to be done about what initially appears to be the underfunding of the TPP in Figure 1.

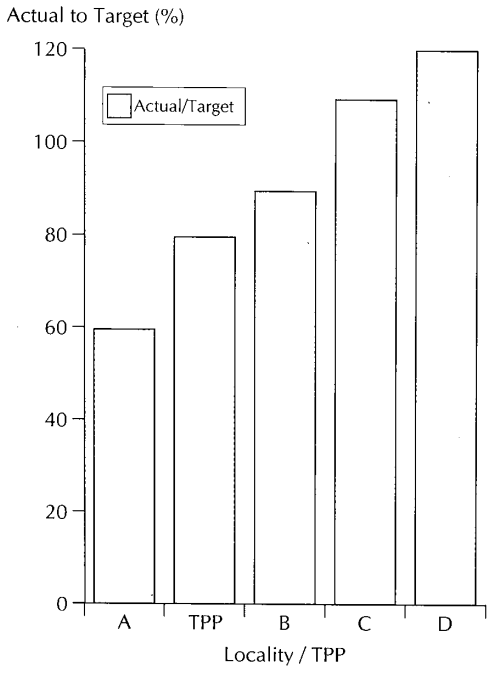


Figure 1: Ratios of actual to target expenditures

Table 1: Resource allocation to authorities and TPPs

Steps in Process	*Authority*	*TPP*
Population	Known from national statistics	Problems over data on lists and overlap with other authorities
Services included	Standard definition	To be determined
Setting per capita rates	Standard formula	Choice of formulae
Historic spend	Known	To be estimated
Pricing services	Average costs + Market Forces Factor	Unclear

Paradox: same objectives different means

Box 2 outlines steps in deriving a capitation target. In the NHS there are now two stages in the allocation of resources: from Government Departments to authorities, and from authorities to TPPs (and localities). The first stage in resource allocation uses standard methods with the considerable advantage of applying consistent principles throughout each UK country at authority level. It is beguiling to extend this approach, which has worked so well at authority level, down to TPPs and localities. This will, however, encounter serious problems. Table 1 compares the steps in the processes of setting budgets for Hospital and Community Health Services (HCHS) by authority and TPP. For the first stage, data on population and past spend are known by authority, there is a standard definition of services, and standard formulae for determining capitation rates and pricing services. For the second stage, there are problems over determining relevant expenditures, populations, and formulae.

Box 2: Five steps in deriving a capitation target

- Define populations
- Define services (e.g. acute, psychiatric, other) for which resources are to be allocated
- Derive per capita rates by risk characteristics (age, sex, other factors) for each service
- Multiply per capita rates by defined populations to derive capitation-based volumes by service
- Use information on prices to derive target expenditure

Although the pilots are called 'total' Purchasers, this is misleading. They are engaged in *selective* not *total* Purchasing. Variations in what pilots do and do not purchase introduce a complex mix across income streams: TPP, standard GP fundholding, and the authority. This variety also means that there is no standard formula, which is available. There are also problems in that populations registered with GPs do not neatly match authority boundaries, and the census areas used as indicators of weights in the national formulae may have patients who register with different practices. These problems are discussed further below.

The main argument of this paper is that although each stage of resource allocation in Table 1 seeks the same objectives, they will best be achieved by developing different approaches:

- In making allocations *to* authorities, standard methods largely enable allocations to be made without entering into negotiations over how to manage the pace of change in allocations[40].

- In making allocations *within* authorities such negotiations will be inevitable. It is thus helpful to have a process, which can contribute to, and illuminate, these negotiations. This paper offers such an approach.

- The national weights may help in setting targets for large localities of 50 000. To make changes, however, it is also likely to be useful to be able to pursue analysis down to practice level. At that level, research suggests that it will be important to analyse variations in medical practice.

3. Methodological problems

Indicators used by authorities in setting budgets

Table 2 gives an overview of methods used in setting TPP budgets from the TP-NET survey of 48 health authorities with TPPs. Replies were received from 36 (75 per cent). Thirteen (36 per cent) of the 36 authorities reported that budgets were based on historic spend only. Out of 23 (64 per cent) authorities where capitation formulae were used in determining budgets, five set budgets based on the results of the formula and eighteen set budgets between historic spending and the results of the formula. As expected, the survey of lead GPs produced similar findings. Eight out of 40 projects reported that budgets were based on historic spend only. Out of 29 projects where capitation formulae were used, twenty-one of the 29 had a budget between historic spending and capitation. Of these, seven had agreed a pace of change for TPP allocations. Hence, for most projects, budgets were set with reference to a capitation formula, and the budget was between historic spending and capitation.

Table 2: Methods of setting budgets as reported by health authorities

Numbers of Authorities	*Past Actual Spend*	*Budget = Target*	*Budget between target & past actual*
13[1]	•		
7		•	
16			•

Note

1 These include two projects, one was allowed to overspend to capitation level without being penalised, and in one capitation was used where data on historic spend were not available

Source: TP-NET 1996 survey of authorities.

Determining past expenditures

Because of problems with lack of data, actual past expenditure by TPP had to be estimated, and deriving these estimates was the most serious and common problem encountered by authorities. This was cited as a contributory factor by each of the 24 authorities that reported delays in determining the TPP allocation. This is because it is difficult to identify activity at practice level, in particular for accident and emergency and community health services.

Each TPP chooses for which subsets of services it will contract. Adjustments have to be made for: services that the TPP has 'blocked back' to the authority (i.e. those services that the TPP are not purchasing directly), spend on fundholding, and those health authority overheads that are still chargeable to the TPP. Eighteen (67 per cent) out of 27 authorities said that there were problems in allowing for costs of services 'blocked back' to the authority. Eight (33 per cent) of the 24 authorities which reported delays in determining the TPP allocation attributed these also to problems in taking account of standard fundholding.

Defining and estimating populations

TPPs consist of general practices, and their lists are the outcomes of patients choosing GPs. This is different from authorities which have their populations defined by administrative areas. In the survey of health authorities 28 reported on the population data they had used. Five (18 per cent) said that they had encountered serious problems in producing the required estimates of the population (total, and distribution by age — or by age and sex). There are two kinds of problems:

- practice populations may be drawn from more than one authority, and
- 'list inflation': that the population registered with a practice includes patents who have died, or moved and babies not yet registered[41].

The survey of authorities showed that eight (28 per cent) out of 28 said that they excluded the population covered by other authorities; that is the TPP acted as insurer for only the majority of their population in one authority. List inflation is a significant problem in cities (in particular London). As few first wave TPPs are in cities, this was not a serious problem for most authorities. Where list inflation was more than 2 per cent, authorities tended take this into account.

Two other kinds of problems can also arise in defining and estimating the population:

- Simply applying a capitation formula to the population registered with a practice will overestimate the likely costs to the practice where patients pay for private care, or are members of the armed forces.
- Patients who are registered with one practice but need hospital treatment in a different area, in particular, students and patients on summer holidays.

Adjustments can be made for members of the armed forces, but it is much more difficult to do this for those who use private health care[42]. The problem posed by University practices is that of a student requiring hospital treatment at home rather than in the area where they are registered. Practices in holiday areas with large numbers of summer visitors pose the opposite problem of admitting patients registered with other GPs (and may also have a transient high-risk population using holiday accommodation at low rents in the winter)[43]. In principle, these problems ought to be taken into account by the invoicing arrangements of the internal market, which ought to identify who is acting as the patient's insurer[44].

Approaches taken by authorities in risk rating TPP populations

At the heart of a capitation formula is the question of what risk rating (if any) to apply to populations to estimate the allocation of resources. This may be considered in terms of three options[45]:

- Crude population only (i.e. numbers of people);
- Take account of risk by age and sex;
- Take account of risk additional to age and sex (described here as an X factor).

The survey of 26 health authorities, which supplied details on the capitation method, used, two used crude population, and the rest weighted the population in various ways. Table 3 gives the various methods used to take account of population needs and prices in the various formulae. Although practice varied, the norm was to:

Table 3: Capitation methods as reported by authorities

Number of Health Authorities	*Age*	*Sex*	*X Factor*
11[1]	National	No	National
3	National	National	National
3	National	Local	National
2	Local	Local	National
2[2]	Local	No	National
2	No	No	No
1	Local	Local	Local
1[3]	National	No	Local
1	National	Local	No

Notes

1 Includes two authorities in which capitation formulae were developed, but budgets were set on historic spend only. Differences in sex examined for one project and not found to be significant.

2 Used national weights for acute, psychiatric services mostly by historic use.

3 Additional needs weights based on SMRs.

Source: TP-NET 1996 survey of authorities.

- use national formulae (with England based on the York formula[46], and Scotland on the SHARE formula[47]) to weight for age and additional needs (referred to here as X factors);
- ignore weightings for sex (in England).

Of the 24 authorities which reported delays in determining the TPP allocation, nine (38 per cent) attributed these to problems in finding an appropriate method to produce a capitation target; and six (25 per cent) to getting data required by the capitation method. More generally, whether allocations were delayed or not, the following problems were identified: 3 (12 per cent) out of 26 encountered serious problems in weighting populations for age; 2 (9 per cent) out of 23 encountered serious problems in applying additional weightings to age (and sex).

There are, of course, two questions about risk rating for TPPs using national weights: are the national weights for risks correct; and, even if they are, will they work for TPPs' practice populations and services?

Do national weights give the correct risk rating for health authorities?

It is impossible to know whether national weights for risks are correct in allocating resources to health authorities. Risk rating belongs to a category of concepts that have been well described as being essentially contested[48]. The root of the problem is that we cannot directly measure the risk in populations for illnesses for which there is effective care. Hence the debate over risk rating is over which proxies are the best measures. As the debate is about proxies it cannot be resolved: if direct measures were available we would use these and not proxies[49].

There are good reasons for taking mortality data as the best proxy measure of illness in populations[50]. But that still poses a host of questions in how to use these data in formulae for allocating resources[51]. Box 3 indicates some of these questions. One of the most serious problems is deciding what weight to give to SMR. The RAWP Report recommended a weight of one, which assumes that a population with a mortality rate 10 per cent higher than expected (based on its age and sex) ought to receive 10 per cent, more resources. The justification for this assumption was that it is the simplest one to make (there being no evidence of what that weight ought to be).

The approach adopted to try to resolve the problems posed in Box 3 has been to undertake empirical analyses to estimate risks from small area variations in utilisation (having taken account of differences in age)[52]. It is indeed likely that variations in risk will produce variations in utilisation. But there are other causes of variations in utilisation:

- at the macro level, by variations in supply, and
- at the micro level by variations in demands by patients and responses by doctors[53].

Box 4 summarises some of the main problems in estimating risk from small area variations in utilisation. There is the further problem that, relying on decennial census data, means that these estimates may become worse over time as characteristics of populations change.

Box 3: Some questions raised by using mortality data as an indicator in formulae for resource allocation

- For which services is mortality a good measure of morbidity: is it only relevant for acute diseases?[1] For what proportion of morbidity is there effective treatment and therefore resources should be made available?
- Mortality is focused on morbidity: to what extent do mortality data capture extra needs for health care of deprived populations?[2]
- Where mortality is a good measure of morbidity, is the SMR an appropriate index?[3]
- If the SMR is an appropriate index, which age range should be used?[4] And if the SMR is based on a restricted age range (under 75), then should it, nonetheless, be applied to the whole population?
- If the SMR is applied to a specified age range, what weight should be used for the SMR?

Notes

1 The purpose of using SMRs is to measure variations in relative need. Bennett and Holland (1977) argued, for example, that there is little geographical variation in the incidence of chronic disease. What matters is whether the residual variation is or is not correlated with the SMR. We do not have data on the incidence of chronic disease. It is likely to be indicated by the census question on long-standing limiting illness. The York Study (Carr-Hill *et al.*, 1994b) found that, in the under-75s, this was strongly correlated with the under-75 SMR (0.81).

2 As the RAWP Report pointed out, SMRs are correlated with all indicators of social deprivation. This has been a consistent finding of subsequent empirical. Carstairs and Morris (1989), for example, found the under-65 SMR to be correlated in Scotland with census variables such as no car (0.74) and unemployment (0.73).

3 The all-age SMR is the most well known index of mortality. There are others: age specific mortality indices, relative mortality index, Yerushalmy's mortality index. Research has shown that use of different indices would significantly change authorities' targets. Mays and Bevan (1987) argued that age specific SMRs would provide more appropriate weighting than all-age SMRs.

4 The all-age SMR is dominated by deaths in the very elderly the under-75 SMR is likely to be more sensitive to variations in underlying risk.

Box 4: Some questions raised by estimating risk from data on utilisation for formulae for resource allocation

Problem	*Outline explanation*	*Effect*
Ecological fallacy	A 'socially-deprived' area may have high utilisation of services, but it may not be the individuals who are 'socially-deprived' in that area who make high use of hospital services.	Is likely to perpetuate unequal existing distribution.
Data quality	The quality of data on utilisation of services at small area level can be poor and data are available for some services only.	Errors of indeterminate impact.
Interaction between risk and utilisation	As Grossman[1] pointed out, there is a two-way interaction between sickness and use of health care: sickness will have an impact on use of hospitals, and vice-versa. This requires simultaneous estimation.[2] The Review of RAWP[3] was methodologically flawed for this reason,[4] but the York study addressed this problem.[5]	This method generates in different models which produce different target allocations for the same authority. A choice has to be made over which model is best.
Spatial interaction between supply and utilisation	It is well known that distance to hospital affects utilisation, and that patients` willingness to travel to hospital is influenced by the existing distribution of hospital services.[6] The York research[7] used a crude spatial interaction model that assumed that people's willingness to travel is the same in rural areas as in cities (which is implausible).	Bias between rural and urban areas.
Medical practice variations and inappropriate treatment	Studies in the UK (and the US) have found correlations between variations in admission rates and generic social indicatorr.[8] In contrast other studies of specific conditions and procedures have revealed variations in rates of treatment which, it is consistently argued, cannot explained by differences in need and are attributed to variations in medical practice.[9] Studies have also found high levels of inappropriate care.[10]	Resources allocated to reflect variations in medical practice rather than capacity to benefit.
Census data	Census data are only available some years after the census and only updated every ten years.	Errors when characteristics of areas change relatively.
Utilisation and social class	The poor, for similar levels of illness, may make less use of health care than the rich do, so on utilisation may underestimate their risk.[11]	Inadequate funding of poor areas.

Notes

1 Grossman M (1972)
2 Use of techniques of simultaneous estimation has been common in US econometric studies which has sought to establish the impact of health care on mortality. See, for example, Auster *et al.* (1969), Hadley (1982).
3 Royston *et al.* (1992).
4 Mays (1989).
5 Carr-Hill *et al.* (1994b).
6 See, for example, London Health Planning Consortium (1979), Taket *et al.* (1986) Biker and Vos (1992).
7 Carr-Hill *et al.* (1994b).
8 See, for example, Royston *et al.* (1992), Carstairs and Morris. (1989), Kirkup and Foster (1990), Knickman and Foltz (1984).
9 The literature on medical practice variation is discussed further below. Papers which specifically focus the problems of using data on utilisation to measure need include: Morgan *et al.* (1987), Sheldon *et al.* (1993), Payne *et al.* (1994), and Bevan (1995).
10 Brook (1994), Brook *et al.* (1990), Chassin *et al.* (1987), Gray *et al.* (1990).
11 Chaturvedi and Ben-Shlomo (1995) present an analysis which suggests that people who are poor have lower surgical rates than would be expected from the rates at which they consult GPs. Newbold *et al.* (1995) show that in Canada income has a significant effect on utilisation of hospital services. They use self-assessments of need for care. The absence of data on sickness in populations ideally required for resource allocation formulae is why the literature is so sparse on what would naively expected to be an obvious empirical question.

Problems in applying national weights to TPPs

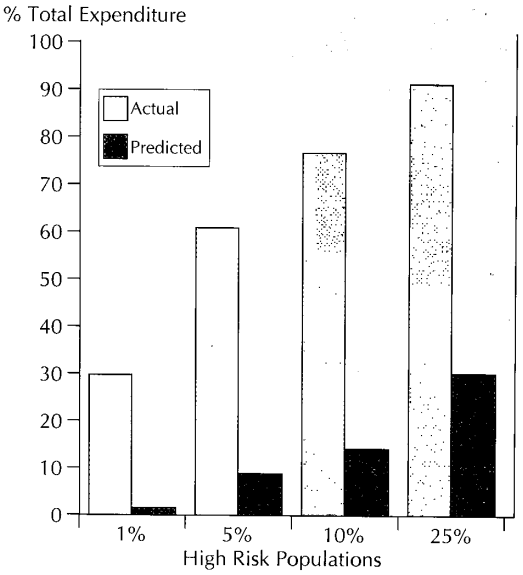
Even if the national weights were correct, there are three additional problems in applying these weights to TPPs:

- The *attribution problem*: when the population in census areas (wards) may be registered in different practices[54].
- The problem of *small* populations that means that formulae will produce estimates surrounded by large errors.
- The problem that *subset of services* included in TPPs may require different risk rating.

When the population in wards may be registered in different practices, authorities have used the obvious method of estimating X factors from national formulae. This is to use census values by wards and weight these by the proportion of each ward in each practice[55]. This assumes that risks in ward populations are attributed at random. This is, however, unlikely. Figure 2 illustrates the skewed distribution of the population risk of using health services using data from the Netherlands[56]. Figure 2 shows that the most expensive one per cent and five per cent of the population account for 30 per cent and 60 per cent of total expenditure on hospital services. Figure 2 also shows that risk rating for age and sex would rate the risk of these groups as about twice as expensive as average, whereas the actual multiple is much higher. Suppose the national methods were right in capturing risk of a ward, and that the same ward (of 5 000 people) is served by more than one practice. If one practice ended up with all the high risk one per cent (50 people), accounting for age and sex would mean that it would receive two per cent of the ward's share of resources, but be faced with demands of 30 per cent of those resources[57].

Where the attribution problem arises, there are two obvious ways of addressing it:

- To develop practice-based data, such as mortality data in the form of Standardised Mortality Ratios (SMRs) as a simpler guide to relative needs of populations of TPPs[58].
- To examine differences between actual expenditures and those indicated by targets[59].



Source: Lamers and van Liet (1995)

Figure 2: Actual and estimated spend on health care for high risk populations

The problem caused by small populations is that the smaller the population, the greater the year-to-year variation in risk[60]. This problem is exacerbated if these populations are insured for a narrow range of services. The impact of year-to-year variations in risk has been examined for all inpatient admissions[61] and for rare costly referrals[62]. Results of each show that these variations can be unmanageably large for annual budgets for populations of 10 000 or less. But, once the budget is extended to about 30 000 person years, the variation is significantly reduced. This does not, however, mean that a TPP (or locality) ought to have to have a population of at least 30 000. A TPP of 10 000 ought to be able to manage year-to-year variations in expenditure, provided that it is allowed a three year period over which to manage these. In turn this means that, for small TPPs, underspends in any given year ought to be carried forward to provide insurance against future overspends.

This suggests that there are likely to be advantages in a two stage approach to resource allocation within health authorities:

- To set cash limits with constraints on annuality of supply at populations of about 50 000; and
- To also calculate indicate targets for resources at smaller levels (say 10 000) and investigate reasons for differences between past spend and future targets at both levels.

TPPs opt for a subset of the services for which authorities contract. If the national weights were correct for all authority services, would they be correct for any subset of services? If 'correct' is taken to mean the same results being derived from relationships between small area variations in utilisation and census data, then the answer will almost certainly be 'No'. We would expect different relationships between risk factors and small area variations of different categories of hospital admissions. This is because variations for some conditions and procedures will mainly reflect underlying sickness, but others will be significantly influenced by variations in medical practice. Research by the York team to develop risk rating for elective hospital admissions found the same set of risk factors and weights could be used as for all non-psychiatric services, but that these ought to be dampened[63].

Accounting for variations in costs

Methods of risk rating are directed at weighting populations for differences in use of volumes of health services. There is a question over how to account for variations in provider prices. Twenty-four authorities supplied details on costs used in capitation formulae for TPPs. Only three (13 per cent) reported using national average prices; normal practice was to use local prices. The most common choice was to use prices paid by the TPP (15, 63 per cent); other local prices used included those paid by the authority, a combination of TPP and authority prices, and regional prices.

The principle underlying the use of capitation targets is that funding insurers by capitation ought to provide individuals of equal risk with equal prospects of treatment. This is not only a question of risk rating but also of ensuring equal purchasing power. This raises the question of how prices paid by the TPP relate to costs used in the capitation formula. Variations can arise from variations in provider costs and in contract currencies.

As mentioned earlier, the problem of variations in provider costs has existed since the RAWP Report. National formulae have sought to compensate for variations in costs that are beyond the control of providers of health care, but not to finance other variations in costs, which would generate perverse incentives towards inefficiency. There is, however, a difference between England and the other countries of the UK in the cost variations in provider costs that are taken into account in national formulae. These reflect characteristics of each country:-

- The formula for England aims to adjust for *market forces*[64].
- Formulae in other UK countries aim to adjust for costs of *sparsity*: the extra travel costs in delivering care to sparsely distributed populations[65].

Where there are significant variations in costs between different providers in an authority, ignoring these will mean different parts will be given different purchasing power. There are two problems here:

- these may be beyond the control of providers: e.g. in capital charges and where one locality uses a London provider;
- these may reflect variations in efficiency of providers.

It is correct for national policy to try to account for the first and ignore the second. It would be mistaken to fund purchasers within an authority in ways that guarantee extra funding for inefficient providers. But, in the short run it may well be necessary to allocate extra money to a locality using a provider with high costs caused by inefficiencies. It would be appropriate to set a pace of change in reducing allocations to take account of a reasonable time for the provider to make its services more efficient.

Another cause of differences in purchasing power is when the TPP and authority use the same provider, but contract in different currencies. If the authority is using a block contract and the TPP cost per case, then actual prices paid are likely to differ, and using the same funding formula would give rise to different purchasing power. The same problem arises when TPPs develop more complex contract currencies with cost bands per day which vary according to how long the patient has been in hospital. There are particular problems when the TPP and authority use different contract currencies and have different age profiles of population.

A problem can also arise even if the TPP and authority use the same provider and the same contract currency, but have different age profiles of

population who are weighted for costs differently from contract prices[66]. A simple example can illustrate this problem. In this simple example a capitation formula divides the population into two groups only: the young and the old. The old are at twice the risk of the young of being admitted to hospital. The cost per admission of the old is 50 per cent higher than the young. The net effect is that the old are funded in the formula at three times the rate of the young. There are four possible scenarios in terms of population mix and contract currencies as illustrated in Box 5.

Box 5: Four scenarios of population mix and contract currencies

Scenario	*Population mix*	*Contract currencies*
A	Same	As in formula[1]
B	Same	Same price for old and young
C	Different	As in formula[1]
D	Different	Same price for old and young

Note
1 This means price per case of the old is 50 per cent higher than for the young.

In scenario A there is no problem: the population mix of the TPP and the authority are the same, and contract currencies are the same as in the capitation formula. In Scenario B, the prices paid by contract are different from those assumed by the formula, but that does not matter because the population mix of old and young are the same. In Scenario C, the population mix of old and young are different, but that again does not matter because, each pays the same prices as they are allowed by the formula. In Scenario D, however, there is a problem because the population mix is different and each pays different prices from those assumed by the formula.

Table 4 illustrates the nature of the problem in scenario D. The TPP has 100 000 of the authority's total population (of 500 000) divided equally between the young and the old. The 400 000 remaining with the authority have seven times as many young as old. Applying the capitation formula to the TPP would mean its total funding would be £200m (£50m for the 50 000 young, and £150m for the 50 000 old). By contracting on a standard cost per admission, it would, however, face costs of only £150m (£50m for the 50 000 young, and £100m for the 50 000 old). The TPP would thus have an extra £50m to spend at the expense of the rest of the authority.

Table 4: Impact of different contract currencies

	Authority (less TPP)	*TPP*
Population (young)	350 000	50 000
Population (old)	50 000	50 000
Risk rating young (£1 000)	£350m	£50m
Prices paid for young	Block	£50m
Risk rating old (£3 000)	£150m	£150m
Prices paid for old	Block	£100m
Deficit / Surplus	- £50m	+ £50m

There is a further problem in taking account of changes over time. Assuming the start point to be an estimate of past spend, revaluing this for inflation will not take account of all the other changes made in an authority's allocation. The aim is to move the TPP's historical share of the authority's resources towards its target share over a number of years. By using historic share of current allocation as the start point means that the TPP is subject to the same pressures as the whole authority.[67]

Local or national formulae

One policy question for authorities is whether they should use local or national formulae in making allocations to TPPs. The UK does not allocate resources for health services to each country using a standard formula[68], and each country of the UK distributes its own allocation using different ways of rating populations for risk and for allowing for differences in provider prices. Each UK country introduced similar capitation formulae in the late 1970s[69] using SMRs as the main X factor with a weighting of one. Since then the English formula has been revised on two occasions to take account of results of empirical analyses of utilisation of health services[70]. Other countries ignored the first analysis but are considering changes in the light of the York study.

There is an obvious attraction within an authority in using the same risk rating to set allocations for TPPs as has been used in determining the authority's allocation. It may be, however, that indicators used

in the national model to discriminate between health authorities do this poorly *within* the same authority[71]. This paper argues, however, that authorities ought not to spend too much effort in seeking their own ways of risk rating given all the formidable problems this implies. It is likely to be more fruitful to focus effort on understanding differences between future targets and past spend. There is likely to always be substantial differences to be investigated, however, sophisticated the formula being used[72].

4. A new approach

Two options

This chapter contrasts two approaches:

- focus on the method to derive the target; or
- accept that the target can be, at best only a rough guide, and
- focus on understanding differences between future target and past allocations.

Focusing on deriving the target

In the first year of budget setting some sought a binding agreement by all parties to the method of deriving the target; that is, that each party was committed in advance to the outcome, whatever changes it generates in previously estimated allocations[73]. When, however, calculations showed that the outcomes of agreed principles was a significant movement of resources, then the losing party sought to abandon the prior agreement. Although this response may seem to be have been an act of bad faith, it is an understandable response by individuals who are bruised and bewildered by the process[74].

The difficulty with any approach based on the assumption that the target is right is that we can be sure that the target will be in error. As this paper has argued, that there are a host of difficulties in risk rating and accounting for costs in working out a future target allocation for a TPP. Where resources are at stake, relationships can easily become antagonistic as the losing party focuses on weaknesses in the method used (of which there will be many).

Any method of resource allocation by capitation can only be a rough guide, but it offers the potential to produce fairer allocations than relying on the haphazard outcomes of the past[75]. As estimates of both future target and past allocations will be subject to errors, it is vital to understand the differences between these two allocations. This, in turn, will show whether changes ought to be made in allocations, and, if so, what responses are required, and over what time scale. It is better to recognise this at the outset than be forced to accept this reluctantly as an admission of failure to produce a perfect target.

Examining differences in price and volume

The first question in examining differences between future target and past allocations is are these caused by differences in:

- prices[76], or
- volumes.

If, for example, a TPP's current allocation exceeds its target, different responses are required if the cause is either high prices or high rates of use of services. Analysis might show that a TPP's current allocation is below its target because of its low prices, and increasing the TPP's allocation would make its budget more inequitable[77].

The nature of finance by capitation is that populations are funded at average rates of treatment by age and sex (and X factors). Hence, if we want to understand why actual rates of admissions are higher or lower than average, we have to investigate reasons for variations from average rates of treatment. This has been the subject of numerous studies that have concluded that the principal explanation is Medical Practice Variation (MPV). This paper outlines here some of the main findings from various studies that have sought to explain variations in volume across populations in terms of MPV.

Volume differences caused by variations in medical practice

MPV exists when different doctors make different decisions about the same or similar patients. The classic example is that of adenoidectomy / tonsillectomy. Bloor[78] offers an entertaining account of various studies of this procedure. He cites the classic study conducted in America before the war (see Box 6). This suggests that you might as well toss a coin as seek a medical opinion on this type of surgery. Pioneering work by Glover also showed large variations in tonsillectomy rates across local

authorities which led him to conclude that this was a 'prophylactic ritual carried out for no particular reason with no particular result'[79]. Since then there has been a vast literature on variations in GP referral rates[80] and hospital admission rates[81]. (Although this is attributed to MPV, variations in demands by patients is likely also to contribute to these variations). Two studies in the US and in England found about 90 per cent of hospital admissions were in a high variation category[82]. These are disturbing results despite concerns over the quality of the data. Further alarm comes from studies by the RAND Corporation that sought to explain variations in admission rates through examination of appropriateness.

Box 6: A classic study of medical practice variation

A study of 1000 New York school children

Sixty-one per cent of these children had already lost their tonsils, the remaining 39 per cent were assessed by a group of school doctors who recommended that 45 per cent of the children should undergo 'T's and A's' and rejected the rest. The *rejected* children were then sent to a second group of doctors who recommended surgery for 46 per cent of them. Those children twice rejected were sent to a third group of doctors who recommended surgery for 44 per cent of them.

Source: This is based on the account by Bloor (1976) of Bakwin (1958).

An intervention is defined to be appropriate where 'the benefits exceed the risk by a wide enough margin to make it worth providing'. Studies found that about half of coronary angiography and coronary artery bypass operations in the Trent region, and 60 per cent of cholecystectomies in the North West Thames Region were performed for inappropriate or equivocal reasons[83].

The key to making progress in the analysis of variations in hospital admissions rates has been clinical information on diagnosis and procedure. Consider, for example, hip fractures and back surgery. Hip fractures are easy to diagnose and the appropriate clinical response is admission to hospital. For back problems, different GPs will have different referral practices, and orthopaedic surgeons will make different decisions on whether or not to operate. It is thus to be expected that the variation observed in admission rates (after standardising for age and sex) is low for hip fractures and high for back surgery. It is, of course, not possible to make sense of these differences by using specialty as a category because they are both orthopaedic procedures.

One of the surprising features of variations in rates of treatment is the way in which massive statistical differences in rates of treatment across areas pass unremarked by patients and doctors in those areas. Thus spend per capita in Boston was about sixty per cent higher than in New Haven (the towns of Harvard and Yale medical schools with similar populations), but doctors who had worked in both towns were unable to assess which had the higher spend[84]. A study examined differences in admission rates by DRG across the towns and found that these were mostly accounted for by medical DRGs known to exhibit high variation, and hence of questionable efficacy (such as treatment for medical back problems)[85].

The normal reaction to redistribution of resources for health care may be caricatured as follows. To cut spend on health care anywhere is to invite a man-made catastrophe. This is because it will directly lead to deaths that are being prevented only by current levels of spending on health care. The accumulating studies of MPV and appropriateness show this caricature to belong in a world of fantasy (that is one without MPV and where all health care is appropriate).

Using information on MPV in resource allocation

Using information from studies of small area variation and knowledge of the type of admission, we can assess whether high rates are justified by need or likely to be manifestations of MPV. If a TPP had higher than expected volumes of admissions (after allowing for age, sex and X factors), the question is in which categories are these high volumes concentrated? Are they, as in Boston, in categories of admission known to be high variation with high levels of inappropriateness (such as tonsillectomy and disc surgery)? Or, are they in

categories of admission known to be low variation and with high levels of appropriateness (such as acute myocardial infarction and hip fractures)? If the higher admission rates are concentrated in high variation categories, this suggests that they may be results of MPV, and that volumes ought to be reduced to average levels. Indeed, redistribution may reduce spending on such types of admission and, in this way, mean that resources for health care are better directed at those risks for which there is effective treatment. If, on the contrary, higher admission rates are concentrated in low variation categories, this suggests that the X factors are inadequately capturing the high risk of this population, and the target is too low.

A study of GP fundholding in Wales found high variation in volumes per capita. Taking account of risk rating (for age and sex alone, and for age and sex and district SMR) did not, however, significantly reduce these variations. And the practice with the highest levels of spend was in a rich retirement area in an authority where other practices were poor former mining villages and towns. Further analysis suggested that variations in volume were caused by differences in patterns of medical practice and not in risk[86].

Box 7 outlines a new approach to setting TPP Budgets[87]. Differences between past actual and future target allocations would be investigated as follows. It would be based on data on inpatients and day cases: classed as Finished Consultant Episodes (FCEs). This is because these data alone have sufficient clinical details to make sense of variations in volume. One way of doing this is to use Healthcare Resource Groups (HRGs)[88] to investigate differences between observed and expected spend by HRG. Are these differences attributable to differences in:

- resource use per FCE (length of stay or cost per day), or,
- volumes of FCEs, and if volumes, whether this appears to be due to
- differences in need or variations in medical practice?

This examination can be undertaken at various levels of population such as TPP (or localities) and practice[89]. This might work as follows[90]. Suppose an authority (of 250 000 population) is divided into 5 localities (of 50 000 population). Using a capitation formula to derive targets gives a start position of ratio of past spend to future targets as given in Figure 1. The focus then is on the two extreme localities A and D. The first question is whether the differences in spend are caused by each paying different prices.

Box 7: A new approach to setting TPP budgets

Key steps

1. Derive local formula based on national advice and guidance but tailored to local circumstances.
2. Calculate target & actual spend by service group (acute psychiatric, acute non-psychiatric etc.)
2. Compare target & actual spend by service group.
3. Examine whether differences caused by price or volume.
4. For volume differences examine whether caused by differences in risk or in medical practice.
5. Negotiate changes in risk rating, provider prices, medical practice.
6. Decide pace of change in allocations.

It may be, for example, that locality A is dominated by a retirement area, and its target allocation is 30 per cent higher than it would be if its population had the same age distribution of the whole authority. Furthermore, the way the locality contracts for services makes no distinction by length of stay, and hence the locality is being allocated in its target an allowance for higher costs per admission which it does not have to pay. Suppose, however, this is shown to only account for 10 per cent of its higher target allocation for its elderly, with 20 per cent for increased volumes. Analysis then focuses on the top 50 HRGs that account for about 50 per cent of admissions and inpatient days. These are all found to have lower than expected numbers of admissions for locality A. In particular, HRGs with high national admission rates for the elderly do not have high admission rates for locality A (e.g. cataracts, knee replacements). There appears to be good reasons for allocating extra resources to this locality to finance increased use of selected services in low variation HRGs deemed to be appropriate.

In contrast, analysis of locality D shows no obvious variation between the price assumed in the target

and the prices paid in contracts, and that most of the top 50 HRGs which account for about 50 per cent of admissions and inpatient days have higher admission rates than are expected. The extra volumes are concentrated in ten HRGs with high variation; five of these are where there are questions of the appropriateness of admission to hospital. These include back problems and dermatological HRGs. Analysis also shows that of the 30 GPs in the locality, three account for half of all referrals to these high variation HRGs.

Given the strong case for increasing funding in locality A, the next step is to work out ways of reducing admissions in locality D. Guidelines are developed on referral and admission for the five most questionable HRGs in the next year. The locality provides feedback to each GP on rates of referral in categories that could result in admission for one of these five HRGs, and to each practice on rates of admission for each HRG. For locality A, collaborative work with public health, GPs in the locality, and hospital consultants explores which admissions might be increased for patients who would benefit from treatment but who are not referred at present. In the next financial year resources are moved from locality D to locality A assuming increases in admissions in D and reductions in A.

This approach needs to be piloted and to show that its benefits outweigh the extra costs of the analytic work it requires. It does, however, offer three advantages over current practices:

- The NHS spends considerable resources in collecting *data* on treatment and costs, but little on translating these *data* into *information* that can be used to improve equity and efficiency of health care resources.
- Without this information, authorities cannot be confident that moving money within an authority will improve either the equity or efficiency with which these scarce resources are used.
- The approach further invites collaboration across providers and purchasers, the different branches of the medical profession (GPs, hospital specialists, and epidemiologists), and authority staff with expertise in finance, information and statistics.

This is not intended to offer a cosy process of maintaining the status quo over the allocation of budgets nor to offer bases for special pleading. The starting point is that applying approximate formulae is likely to reveal massive variations in per capita spend which are indications that resources are being used inequitably and inefficiently. The purpose of providing extra information is not to undermine the case for redistribution, but to strengthen this by showing what actions need to be taken for this to happen[91].

Implications for the NHS

In an entertaining essay, Evans[92] considered the paradox of why, given twenty years of research into medical practice variation, the consistent findings from various countries had been so consistently ignored by those responsible for shaping policies for health services. His answer was consisted of two explanations: that addressing medical practice variations meant moving onto the treacherous ground of clinical autonomy[93]; and, that Governments have discovered that they can contain the costs of health care whilst allowing MPV to flourish unchallenged.

Evans' argument is that it has been possible to contain costs nationally and allow local variations in medical practice. Applying capitation formulae to localities will, however, reveal variations in prices and volumes of admissions and put analysis of variations in prices and in medical practice at the heart of budget setting. To make sense of these variations requires a multi-disciplinary approach.

The impression, gained from the national evaluation of TP, is that often only the lead GP in the TPP is involved in problems of resources. Resource allocation for locality purchasing ought to be developed through a multi-disciplinary group with authority staff (e.g., finance, public health, and statistics and information), a number of GPs and hospital doctors[94]. GPs who have become fundholders say that through the information they receive, they have gained a much broader knowledge of the services their patients receive than was ever possible from the limited and piecemeal data they received before. The information fundholders receive is, however, localised to their practices. (Unless effort is made to provide inter-practice comparisons.) An advantage of the approach proposed here is that it offers a framework for comparative information across practices.

This paper has tried to describe why there can properly be different methods for resource allocation at different levels. It is reasonable at national level to use a standard formula and over time phase actual allocations towards targets[95]. It also makes sense to do this using standard criteria. It would be hopeless for the NHSE to try to assess the pace of change by authority in terms of the particular actions each authority would have to take to achieve target levels. But, within authorities, an attempt to apply the same approach is fraught with difficulties. Hence this paper has argued for analysis of local variations in price and in volume. Despite all the difficulties entailed by the approach outlined here, I believe this offers potential for the first major improvement in the way in which resources are allocated in the NHS since the RAWP Report of 1976.

Notes

1 Secretaries of State for Health for Health, Wales, Northern Ireland and Scotland (1989).

2 Note, for economy, in the rest of this paper, *authority* is taken also to apply to Health Boards in Scotland, and Health and Social Services Boards in Northern Ireland.

3 Prescribing and some staff costs in primary care; and outpatients, elective inpatients and diagnostic costs in hospital care.

4 Audit Commission (1996).

5 Mays *et al.* (1996).

6 Secretaries of State for Health for Health, Wales, Northern Ireland and Scotland (1989).

7 The NHS Executive (1996) has since issued guidance on developing capitation benchmarks.

8 See Dixon (1994) and Dixon *et al.* (1994).

9 This depended on the cause of these savings. Were they, for example, because the budget had been set too generously, or through windfall gains, or through GPs substituting their services for use of outpatients, or through choice of a more efficient provider, or through providers shifting costs from contracts with GPFHs to the authority? Reducing future budgets for GPFHs because they make savings obviously then reduces their incentives to make savings in future.

10 For GPFHs, guidance issued in June 1996 required offers for all components of the budget to be made by 14 February, 1997 for the financial year 1997–98 (NHS Executive, 1996a).

11 For TPPs this implies the authority being in dispute with itself, since each TPP is formally a sub-committee of its authority.

12 For example, New Zealand, see Malcolm (1997).

13 Roper (1988) then Administrator of the Health Care Financing Administration foresaw this after five years experience of the Prospective Payment System by Diagnosis-Related Group (DRG). Reinhardt (1997) gives a recent account of developments in the US and shows how prescient Roper was.

14 It was no part of the model of the internal market proposed by Enthoven (1985) which was firmly based on that of the US HMO, with GPs employed by health authorities as gatekeepers.

15 This used to be called the 'efficiency trap' before the introduction of the internal market. The problem remains but has been renamed. One solution to the problem of 'overperformance' would be to introduce volume-based methods of funding in which money really did follow the patient. But that would threaten benefits of the current system which cash limits total spend and distributes this equitably according to estimated relative needs of populations. One way of addressing this problem is a mixed system in which part of the costs of care are paid according to the volumes treated and part through capitation. This is being tried in Norway.

16 See Mays and Dixon (1996).

17 Bevan *et al.* (1980).

18 This continues to be done by Medical Practices Committees exercising negative direction: refusing applications for positions in 'over-doctored' areas.

19 Before the RAWP Report, some attempt was made by the Crossman formula, but that had little impact. See Mays and Bevan (1987).

20 This implemented a Labour Party manifesto commitment of the 1970s to territorial justice. The intellectual case for territorial justice have been developed by Bleddyn Davies – I am grateful to Brian Abel-Smith for pointing this out.

21 Bevan *et al.* (1980).

22 In particular the study for the Review of the Resource Allocation Working Party Formula Department of Health and Social Security (1988).

23 Crombie and Fleming (1988)

24 DHSS (1976).

25 See, for example, Townsend (1987) and Sheldon *et al.* (1993).

26 See, for example, Williams (1975).

27 See Watt *et al.* (1994) for a discussion of health and health care in rural areas.

28 See Bennett and Holland (1997), Sheldon *et al.* (1994), Martin and Smith (1995).

29 A finding recently confirmed by Ben-Shlomo *et al.* (1996).

30 See, for example, Davey-Smith *et al.* (1990), Davey-Smith and Eggar (1993).

31 See, for example, Coast *et al.* (1997).

32 Callingham (1997).

33 Bevan (1995).

34 Carr-Hill *et al.* (1994a, 1994b), Smith *et al.* (1994), Sheldon *et al.* (1994)

35 See Bloor *et al.* (1978)

36 Culyer (1976), Williams (1978), Coast *et al.* (1997) discuss need in terms of capacity to benefit and these arguments all apply to risks as defined in this paper. Measuring risks without regard to these facts may be acceptable where individuals choose to pay for this through private health insurance and are content with paying premiums for ineffective care. But the NHS is a health insurance system funded by taxation. In this system, resources ought to be targeted on effective care (allowing individuals to take out extra finance for other types of care if they want).

37 In principle, the internal market offered the potential to do this with purchasers, as insurers, are funded equitably, who then use the market to choose between competing providers to promote efficiency. In practice, of course, the market was much more complicated!

38 Eight (80%) said this was because the target allocation was unacceptable to the authority (because the TPP's target was much higher than its estimated actual expenditure). Two (20%) said this was because the target allocation was unacceptable to the TPP (because the TPP's target was much lower than its estimated actual expenditure).

39 It is also important to manage allocations for small populations over a number of years to even out year-to-year variations in risk (see below).

40 The Department of Health in Scotland faces the problems of making allocations to small populations of Island health boards which are faced by health authorities only in the rest of the UK.

41 List inflation is indicated by comparing authority estimates of populations registered with GPs with estimates from censuses. This gives discrepancies at authority, but not practice level.

42 This problem particularly applies in setting budgets for GPFHs. This is because individuals who opt for private health care are likely to do this to avoid (or reduce) waiting times for outpatients and elective inpatient care; and these are the services covered by standard GP fundholding. This raises problems in principle and in availability of data. There is an issue of principle over whether the allocation of public money for health care within authorities ought to be adjusted according to local use of private services: this is not done in making national allocations of resources to authorities for health care – or education. Data on private health insurance and use of private health care are not publicly available in a way that can be used to modify budgets. This could be remedied by requiring all providers of hospital services to supply data on details of all treatment to the individual's authority of residence (as is required by statute in many states in the US).

43 These problems also arise in applying national formulae to University cities and holiday towns but is more serious proportionately at practice level. There is also a problem in risk rating students, as they are likely to be a different risk (probably low risk) from the general population of that age group. This is an example of the general problem of estimating risk factors additional to age (and sex) at practice level which is discussed below.

44 Indeed that seems to be the only way of dealing with this problem. But current arrangements are obviously not as watertight as we understand practice to be in the US. To make arrangements in the NHS the same as in the US would involve more attention to invoicing. This runs counter to the spirit of new policies that are intended to reduce transaction costs of the internal market (NHS Executive, 1997a).

45 I am grateful to Stephen Birch for this helpful presentation.

46 At the time of this survey, the English formula used different weights for acute and psychiatric services based on results of the York Study: see Carr Hill *et al.* (1994a, and 1994b) Smith *et al.* (1994). At that time the formula for other services was based on crude population only. Since then weights for community health services have been introduced based on the study of Buckingham *et al.* (1996). For a summary of the current position, see NHS Executive (1997b).

47 At the time of this survey, the Scottish formula was that of the SHARE Report, see Scottish Home and Health Department (1977). Since then, some are using the 'Hancock' formula.

48 That is a concept whose application is inherently a matter of dispute. Gallie (1955–6) quoted by Lukes (1974).

49 This is not merely the lack of routine collect data on illness. Even if such data were available, they alone would be insufficient for resource allocation. We ideally require in addition to data on illness, an estimate of the proportion capable of benefiting from treatment and how much funding ought to be made available for different types of treatments. Given what we know about variations in medical practice and in hospital costs, we do not want systems that fund these variations without challenging them. But the information we require to determine idealised targets to do this in one step is not, and never will be, available. Hence, it is impossible to develop such a formula from the bottom up. This paper indicates how challenges can be made to variations in medical practice and hospital prices by comparing top-down target allocations with actual expenditure derived from the bottom up.

50 Mays and Bevan (1987), Sheldon *et al.* (1993).

51 This account draws on work with Trevor Sheldon for the Northern Ireland Department of Health and Social Services.

52 Mays (1995).

53 These micro-variations are described as variations in medical practice reflecting the crucial role of doctors as agents for patients.

54 This problem does not arise where practices have a monopoly across wards (or where localities are defined by wards rather than by practices).

55 This is how Jarman scores are calculated for deprivation payments to GPs.

56 See Lamers and van Vliet (1995). The finding of the skewed distribution of expenditure is common. Scheffler (1989) reported US data showing that, the ranking the elderly by spend, the top 7.7% of the accounted for over 71% of total expenditure. Matsaganis and Glennerster (1994) reported that the top 5% in a fundholding practice accounted for 68.4% of total expenditure.

57 The normal problems associated with the skewed distribution of expenditures shown in Figure 2 are that of 'cream-skimming': i.e. insurers will try to exclude the high-risk population. It has been argued, however, that it is very difficult to identify who these people are, and hence there are limits to what those who aim to practice cream skimming can achieve. The problem, which Scheffler (1989) identified, is rather different: that practices may end up with a high-risk population which is not identified by routinely reported data: e.g., using national methods of risk rating by ward. He alleged that this would be the Achilles heel of GP budget holding. Scheffler argued that the problem caused by a high-risk population would be acute with small populations, and implied that using larger populations would solve the problem. It does mean that a large population is less likely to be vulnerable by including a high proportion of the high risk population by chance. But it does not resolve the problem of biased selection, if, for example, a large TPP includes a number of practices that tend to attract the high risk patients. See Scheffler (1989), Matsaganis and Glennerster (1994), van de Ven (1994).

58 But the problems that are indicated by Box 3 in using SMRs ought to be addressed.

59 The point here is that the problem is caused by a small number of individuals with high expenditures. Figure 2 shows that trying to identify these people from routine data on the whole population is rather like looking for needles in a haystack: 80 per cent of the population make negligible use of health care. But by identifying actual expenditures the individuals incurring high spend may be identified and these expenditures may be examined to see whether they are justified. Examining actual expenditures also offers a way of checking whether what drives increased allocations within authorities is reflected in where the money is actually spent. For example, retirement areas will attract high levels of funding because of the high proportion of elderly: but is this extra funding actually spent on the elderly?

60 The NHS Executive is to issue guidance on population size later this year.

61 Martin *et al.* (1997).

62 Bachmann and Bevan (1996).

63 This included data on waiting times. Martin and Smith (1995). The NHS Executive has concluded this shows that the non-psychiatric acute index of risk factors is a 'robust' indicator of the relative need for GPFH inpatient procedures NHS Executive (1996a). This work showed, however, that the weighting for risk for these factors was lower than for emergency admissions. Hence estimating risk from data on utilisation shows that subsets of acute admissions will produce different weighting for risk. As TPPs opt for different sets of services, there is no easy way of using the national set of weights on a consistent basis for TPPs.

64 The RAWP report allowed for higher pay in London through London Weighting (DHSS, 1976), subsequently a Market Forces Factor (MFF) was introduced DHSS (1980). The current MFF aims to take account of geographical variations in pay, land values and building costs. This was criticised by the Health Committee of the House of Commons (1996). In reply, the Secretary of State for Health (1996) stated that the results of the work on Market Forces were being evaluated.

65 This is for community health services and ambulances only.

66 I am grateful to Brian Maynard-Potts for explaining this to me.

67 I am again grateful to Brian Maynard-Potts for pointing this out.

68 In 1995–96, for example, spend per capita on health and personal social services was £806 in England, £989 in Scotland, £917 in Wales, and £924 in Northern Ireland (HM Treasury, 1997).

69 Bevan *et al.* (1980).

70 Mays (1995).

71 The numbers of New Commonwealth Immigrants, for example, acts as a risk factor for psychiatric services in the national formula in England and discriminates between inner cities and rural areas. Within a rural area other indicators might be required to discriminate between different populations.

72 Remember the objective is not to explain all the variation as this would merely replicate the current pattern of spend! The complex models developed by the York team (Carr Hill *et al.*, 1994b) only explained about half the total variation in utilisation of *volumes* of services. Allowing for variations in provider *costs* is likely to result in extra variation in spend per capita within authorities.

73 The guiding principle here is that advocated by Rawls (1971) of resolving distributional issues in a society using a 'veil of ignorance': e.g. each individual decides what a labourer and doctor ought to be paid without knowing which job he or she will do when the veil is lifted.

74 Most people seem to believe that any objective assessment of their needs for resources for health care could only produce one outcome – to show that they have been underfunded in the past.

75 The introduction of a formula based on capitation is almost certainly a better guide to future allocations than past levels of expenditures. But once a capitation formula is in place, it is much more difficult to be sure that future refinements do produce more equitable allocations.

76 A study of fundholding allocations in Wales found that one practice was paying prices on average about 30% above the average.

77 This could happen for two reasons. One is that the TPP's providers charge low prices because they have low capital charges. The other is that the TPP is in a retirement area with high numbers of elderly, and has a target which assumes higher costs per admission for elderly patients, but the TPP actually pays an average price for all admissions (the example given above in Table 4).

78 Bloor (1976).

79 Glover (1938).

80 Wilkin and Dornan (1990).

81 Andersen and Mooney (1987), Sanders *et al.* (1989), Paul Shaheen *et al.* (1987).

Payne *et al.* (1994).

82 Wennberg *et al.* (1984) analysed hospital market areas in Maine, and McPherson *et al.* (1996) analysed district health authorities in four regions in England (accounting for two-thirds of the population).

83 Brook (1994).

84 Wennberg (1985).

85 Wennberg *et al.* (1987).

86 Bevan G, Sheldon T. (1996)

87 I am grateful to Karen Hancock for this formulation.

88 These are modelled on Diagnosis-Related Groups (DRGs) but have been designed for English classifications of procedures. See Sanderson *et al.* (1995).

89 It is worth reporting observed (but not expected) values by individual GP.

90 I am grateful for discussion with Mark Callingham over the advantages of risk rating at the level of 50 000 population of England.

91 This analysis may also lead to a more informed debate about what future spend on the NHS ought to be. It could translate to the level of practices how changes in total spend affect services at a scale we can understand.

92 Evans (1987).

93 As Evans graphically put it: 'The first man over the barricades gets the spear through his chest'!

94 It may, for example, suggest that GP referrals to hospitals be agreed through a process of peer review. It may also suggest GPs and hospital doctors collaborating to understand variations in admission rates and in this way reduce levels of inappropriate care.

95 Although this approach has to be modified in Scotland for the small island health boards.

Bibliography

Andersen TV, Mooney G (1987) *The Challenges of Medical Practice Variation*, London: MacMillan.

Auster R, Leveson I, Sarachek D (1969) The production of health, an exploratory study. *Journal of Humn Resources* **4**: 411–436.

Bachmann M, Bevan G (1996) Determining the size of a total purchasing site to manage the financial risks of rare costly referrals *BMJ*, **313**: 1054–57.

Audit Commission (1996) *What the Doctor Ordered.* London: HMSO.

Bennett AE, Holland WW (1977) Rational planning or muddling through? *Lancet, i*: 464–466.

Bevan G (1995) Ways of seeing: Explaining variations in use of acute hospital resources. *Int J Epidemiol* **24**: S103–S108.

Bakwin H (1958) The tonsil-adenoidectomy enigma. *J. Pediat* **53**: 339–361.

Ben-Shlomo Y, White IR, Marmot M. (1996) Does the variation in the socioeconomic characteristics of an area affect mortality? *BMJ* **312**: 1013–14.

Bevan G, Copeman H, Perrin J, Rosser R (1980) *Health Care Priorities and Management.* London: Croom Helm.

Bevan G, Sheldon T (1996) Review of methods of risk rating in the UK National Health Service. In *Risk Structure Compensation.* Proceedings of an AIM workshop, Maastricht, 11 October 1995. Brussels: Alliance Nationale des Mutualites Chretiennes, 1996, 41–56.

Biker JA, de Vos AF (1992) A regional supply and demand model for inpatient hospital care. *Environ Planning A*, **24**: 1097–1116.

Bloor M (1976) Bishop Berkeley and the adenoidectomy enigma: an exploration of variation in the social construction of medical disposals. *Sociology* **10**: 43–61.

Bloor MJ, Venters GA, Samphier ML (1978) Geographic variation in the incidence of operations on the tonsils and adenoids: an epidemiological and sociological investigation (Parts 1 and 2). *Journal of Laryngology and Otology* **XCII**: 791–801 and 883–895.

Brook RH (1994) Appropriateness: the next frontier. *BMJ* **308**: 218–19.

Brook RH, Kossekoff JB, Park RE *et al.* (1990) Diagnosis and treatment of coronary disease: comparison of doctors' attitudes in the USA and the UK. *Lancet* **I**: 750–53.

Buckingham K, Bebbington A, Campbell S *et al.* (1996) *Interim Needs Indicators for Community Health Services.* Canterbury: Personal Social Services Research Unit, University of Kent at Canterbury.

Buxton MJ, Klein RE (1978) *Allocating health resources: A commentary on the Report of the Resource Allocation Working Party.* (Royal Commission on the National Health Service, Research Paper 2) London: HMSO.

Callingham M (Chair) (1997) *Sub-Health Authority Resource Allocation Working Group: Technical Target Setting Issues.* Bristol: NHS Executive, South and West.

Carr Hill RA, Sheldon TA, Smith P *et al.* (1994a) Allocating resources to health authorities: development of method for small area analysis of use of inpatient services. *BMJ* **309**: 1046–1049.

Carr Hill RA, Hardman G, Martin S *et al.* (1994b) *A Formula for Distributing NHS Revenues Based on Small Area Use of Hospital Beds* University of York.

Carstairs V, Morris R (1989) Deprivation, mortality and resource use. *Community Med* **11**: 364–72.

Chassin MR, Kossekoff J, Park RE *et al.* (1987) Does inappropriate use explain geographic variations in the use of health care services? *JAMA* **258**: 2533–37.

Chaturvedi N, Ben-Shlomo Y (1995) From the surgery to the surgeon: does deprivation influence consultation and operation rates? *Br J Gen Pract* **45**: 127–31.

Coast J, Bevan G, Frankel S (1997) An equitable basis for priority setting? In Coast J, Donovan J, Frankel S. *Priority Setting: The Health Debate*. Chichester: Wiley, 141–166.

Crombie DL and Fleming DM (1988) General Practitioner referrals to hospital: the financial implications of variability. *Health Trends* **20**: 53–56.

Culyer AJ (1976) *Need and the National Health Service*. Oxford: Martin Robertson.

Davey-Smith G, Bartley M, Blane D (1990) The Black Report on socioeconomic inequalities in health 10 years on. *BMJ* **301**: 373–377.

Davey-Smith G, Egger M (1993) Socioeconomic differentials in health and wealth. *BMJ* **307**: 1085–86.

Department of Health and Social Security (1976) *Sharing Resources for Health in England. Report of the Resource Allocation Working Party* (The RAWP Report). London: HMSO.

Department of Health and Social Security (1980) *Report of the Advisory Group on Resource Allocation (The AGRA Report)*. London: DHSS.

Department of Health and Social Security (1986) *Review of the Resource Allocation Working Party Formula (Report by the NHS Management Board)*. London: DHSS.

Department of Health and Social Security (1988) *Review of the Resource Allocation Working Party Formula (Final Report by the NHS Management Board)*. London: DHSS.

Dixon J (1994) Can there be fair funding for fundholding practices? *BMJ* **308**: 772–775.

Dixon J, Dinwoodie M, Hodson D *et al.* (1994) Distribution of NHS funds between fundholding and non-fundholding practices. *BMJ* **309**: 30–34.

Evans RG (1987) The dog in the night time. In: Andersen TV, Mooney G (1987) *The Challenges of Medical Practice Variation*, London: MacMillan, 117–52.

Enthoven AC (1985) *Reflections on the Management of the NHS*. London: Nuffield Provincial Hospitals Trust.

Gallie WB (1955–6) Essentially contested concepts. *Proceedings of the Aristotelian Society* **56**: 167–98.

Glover AJ (1938) The incidence of tonsillectomy in schoolchildren. *Proceedings of the Royal Society of Medicine* **31**: 1219–1236.

Gray D, Hampton JR, Bernstein SJ, Brook RH (1990) Audit of coronary angiography and by pass surgery. *Lancet* **335**: 1317–20.

Grossman M (1972) *The Demand for Health: A Theoretical and Empirical Investigation*. New York: Columbia University Press.

Hadley J (1982) *More Medical Care, Better Health*. Washington: Urban Institute Press.

Health Committee of the House of Commons (1996) *Allocation of Resources to Health Authorities*. Second Report, Volumes 1 and 2, (HC 477-I, 477-II). London: HMSO.

HM Treasury (1997) *Public Expenditure*. Cm 3601. London: HMSO.

Judge K, Mays N (1994) Allocating resources for health and social care in England. *BMJ* **308:** 1363–6.

Kirkup B, Foster D (1990) How will health needs be measured by districts? Implications of variations in hospital use. *J Publ Hlth Med:* **12**: 45–50.

Knickman JR, Foltz A-M (1984) Regional differences in hospital utilisation. How much can be traced to population differences? *Medical Care* **22**: 971–986.

Lamers L, van Vliet R (1995) Multi-year diagnostic information from prior hospitalisation as a risk adjuster for capitation payments. Paper presented at the third European Conference on Health Economics, Stockholm, Sweden, August 1995.

Leape LL, Park RE, Solomon DH *et al.* (1990) Does inappropriate use explain small area variations in the use of health care services? *JAMA* **263**: 669–72.

London Health Planning Consortium (1979) *Acute Hospitals in London: A Profile by the London Health Planning Consortium*. London HMSO.

Lukes S (1974) *Power*. London: Macmillan.

McPherson K, Downing A, Buirski D (1996) *Systematic Variation in Surgical Procedures and Hospital Admission Rates*. London: London School of Hygiene and Tropical Medicine, PHP Departmental Publication No. 23.

Malcolm L (1997) GP budget holding in New Zealand. *BMJ* **314**: 1890–92.

Martin S, Smith P (1995) *Modelling Waiting Times for Elective Surgery*. York: Centre for Health Economics, University of York.

Martin S, Rice N, Smith PC (1997) *Risk and the GP Budget Holder*. York: Centre for Health Economics, Discussion Paper 153.

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

Mays N (1989) NHS resource allocation after the 1989 white paper: a critique of the research for the RAWP review. *Community Med* **11**: 173–86.

Mays N (1995) Geographical resource allocation in the English National Health Service, 1971–1994: The tension between normative and empirical approaches. *Int J Epidemiol* **24**: S96–102.

Mays N, Bevan G (1987) *Resource Allocation in the Health Service*. London: Bedford Square Press

Mays N, Dixon J (1996) *Purchaser Plurality in UK Health Care*. London: King's Fund Publishing.

Morgan M, Mays N, Holland W (1987) Can hospital use be a measure of need for health care? *J. Epidemiol Comm Hlth*, **41**: 269–74.

Mays N, Goodwin N, Bevan G, Wyke S (1996) *Total Purchasing: A profile of the national pilot projects*. London: King's Fund Publishing.

NHS Executive (1996a) *Progress Report of the Resource Allocation Group*. Leeds: NHS Executive.

NHS Executive (1996b) *General Practitioner Fundholder Budget-Setting: The National Framework*. Leeds: NHS Executive EL(96)55.

NHS Executive (1997a) *Changing the Internal Market*. Leeds: NHS Executive EL(97)33.

NHS Executive (1997b) *HCHS Revenue Resource Allocation to Health Authorities: Weighted Capitation Formula*. Leeds: Resource Allocation and Funding Team, NHS Executive.

Paul Shaheen P, Clark JD, Williams D (1987) Small area analysis: a review and analysis of the North American literature. *J Health Polit Policy Law* **12**: 714–809.

Payne JN , Coy J, Patterson S, Milner PC (1994) Is use of hospital services a proxy for morbidity? A small area comparison of the prevalence of arthritis, depression, dyspepsia, obesity and respiratory disease with inpatient admission rates for these disorders in England. *J. Epidemiol Comm Hlth* **48**: 74–78.

Rawls J (1971) *A Theory of Justice*. Oxford: Oxford University Press.

Reinhardt UE (1997) A social contract for 21st century health care: three-tier health care with bounty hunters. *Health Economics* **5**: 479–499.

Roper WL (1988) Perspectives on physician payment reform: the resource-based relative-value scale in context. *N Eng J Med* **319**: 865–67.

Royston GHD, Hurst JW, Lister EG, Stewart PA (1992) Modelling the use of health services by populations of small areas to inform the allocation of central resources to larger regions. *Socio-economic planning sciences* **26**: 169–80.

Sanders D, Coulter A, McPherson K (1989) *Variations in hospital admission rates: a review of the literature*, London: King Edward's Hospital Fund for London.

Sanderson H, Anthony P, Mountney LM (1995) Healthcare Resource Groups. Version 2. *Public Hlth Med* **17**: 349–54.

Scottish Home and Health Department (1977) *Report of the Working Party on Revenue Resource Allocation*. Edinburgh: HMSO.

Scheffler R (1989) Adverse selection: the Achilles heel of the NHS reforms *Lancet* **i**: 950–952.

Secretary of State for Health (1996) *Allocation of Resources to Health Authorities*. Cm 3394. London: HMSO.

Secretaries of State for Health, Wales, Northern Ireland and Scotland (1989).*Working for Patients*. CM555. London: HMSO.

Sheldon TA, Smith GD, Bevan G (1993) Weighting in the dark: resource allocation in the new NHS. *BMJ*, **306**: 835–839.

Sheldon TA, Smith P, Borowitz M *et al.* (1994) Attempt at deriving a formula for setting general practitioner fundholding budgets. *BMJ* **309**: 1059–1064.

Smith P, Sheldon TA, Carr Hill RA *et al.* (1994) Allocating resources to health authorities: results and policy implications of small area analysis of use of inpatient services. *BMJ* **309**: 1049–1054.

Taket AR, Mayhew LD, Gibberd RW *et al. (1986)* RAMOS: a model of the spatial allocation of health care resources. In World Health Organization *Health Projections in Europe*. Copenhagen: WHO Regional Office for Europe, 218–236.

Townsend P (1987) Deprivation. *Jnl Soc Policy* **2**: 127–146.

Watt I, Franks A, Sheldon TA (1994) Health and health care of rural populations in the UK: better or worse? *J Epidem Commty Health* **48**: 16–21.

Wennberg JE, McPherson K, Caper P (1984) Will payment based on diagnosis-related groups control hospital costs? *New Eng J Med* **311**: 295–300.

Wennberg JE (1985) On patient need, equity, supplier-induced demand, and the need to assess the outcome of common surgical practices. *Medicare* **23**: 512–520.

Wennberg JE (1987) The paradox of appropriate care. *JAMA* **258**: 2568–69.

Wennberg JE, Freeman JL, Culp WJ (1987) Are hospital services rationed in New Haven or over-utilised in Boston? *Lancet* **i**: 1185–1189.

Williams A (1978) 'Need' – an economic exegesis. In Culyer AJ, Wright KG (eds) *Economic Aspects of Health Services*. Oxford: Martin Robertson, 32–45.

Williams R (1975) *The Country and the City*. St Albans: Paladin.

Wilkin D, Dornan C (1990) *General Practitioner Referrals to Hospital. A review of research and its implications for policy and practice*. Manchester: Centre for Primary Care Research, University of Manchester.

Total Purchasing National Evaluation Team (TP-NET)

The evaluation is led by Nicholas Mays, Director of Health Services Research at the King's Fund Policy Institute, London. The different consortium members are listed below, together with their research responsibilities.

KING'S FUND POLICY INSTITUTE
11–13 Cavendish Square, London, W1M 0AN
T: 0171 307 2400 F: 0171 307 2807

Lead: Nicholas Mays
Other members: Nick Goodwin, Gill Malbon, Julian Le Grand, Jennifer Dixon, Amanda Killoran, Jo Ann Mulligan

Project Responsibilities: Hertford, Hemel Hempstead, Hillingdon, New River, St Albans, Stevenage, Attleborough, South Bucks, Belper, Keyworth, Long Eaton, Melton Mowbray.

Other Main Responsibilities: Process evaluation co-ordination (Mays, Goodwin); A&E services and emergency admissions (Dixon, Mulligan); monitoring at all TPPs (Malbon); case studies (Mays, Goodwin, Killoran, Malbon).

NATIONAL PRIMARY CARE R&D CENTRE
Manchester: University of Manchester, 5th Floor, Williamson Building, Oxford Road, Manchester, M13 9PL
T: 0161 275 7600 F: 0161 275 7601
Salford: PHRRC, University of Salford, Davenport House, 4th Floor, Hulme Place, The Crescent, Salford, M5 4QA
T: 0161 743 0023 F: 0161 743 1173
York: YHEC, University of York, YO15 4DD
T: 01904 433620 F: 01904 433628
CHE, University of York, York, YO1 5DD
T: 01904 433669 F: 01904 433644

Leads: Brenda Leese (Manchester and CHE), Linda Gask (Manchester), Jennie Popay (Salford), John Posnett (YHEC)
Other members: Ann Mahon, Martin Roland, Stuart Donnan, John Lee, Andrew Street

Project Responsibilities: High Peak, North Lincolnshire, Rotherham, Sheffield South, Ellesmere Port, Knutsford, Liverpool Neighbourhood, Newton le Willows, Wilmslow, Ribblesdale, Southbank, North Bradford, Wakefield, York.

Other Main Responsibilities: Transaction costs (Posnett and Street); service provision for the seriously mentally ill (Gask, Roland, Donnan and Lee); service provision for people with complex needs for community care services (Popay); relations with health authorities (Leese and Mahon); maternity (Posnett).

DEPARTMENT OF SOCIAL MEDICINE, UNIVERSITY OF BRISTOL
Canynge Hall, Whiteladies Road, Bristol, BS8 2PR
T: 0117 928 7348 F: 0117 928 7339

Lead: Kate Baxter
Other members: Max Bachman, Helen Stoddart

Project Responsibilities: Bewdley, Birmingham, Bridgnorth, Coventry, Solihull, Worcester, Saltash, South West Devon, Thatcham.

Other Main Responsibilities: Budgetary management (Baxter); risk management (Bachman); use of evidence in purchasing (Stoddart); case studies (Baxter).

DEPARTMENT OF GENERAL PRACTICE, UNIVERSITY OF EDINBURGH
20 West Richmond Street, Edinburgh, EH8 9DX
T: 0131 650 2680 F: 0131 650 2681

Leads: Sally Wyke
Other members: Judith Scott, John Howie, Susan Myles,

Project Responsibilities: Durham, Newcastle, Tynedale, Aberdeen West, Ardersier & Nairn, Grampian Counties, Lothian, Strathkelvin

Other Main Responsibilities: Maternity (Wyke); monitoring of participants' views (Wyke); prescribing (Howie); community care (Wyke and Scott, formerly Myles); case studies (Evans)

INSTITUTE FOR HEALTH POLICY STUDIES, UNIVERSITY OF SOUTHAMPTON
129 University Road, Highfield, Southampton, SO17 1BJ
T: 01703 593176 F: 01703 593177

Lead: Ray Robinson
Other members: Philippa Hayter, Judy Robison, David Evans

Project Responsibilities: Dorset, Romsey, Trowbridge Bath & Frome, Winchester, Bexhill, East Grinstead, Epsom, Kingston & Richmond, Merton Sutton & Wandsworth, West Byfleet.

Other Main Responsibilities: Contracting methods (Robinson, Raftery and Robison).

HEALTH ECONOMICS FACILITY, HSMC, UNIVERSITY OF BIRMINGHAM
40 Edgbaston Park Road, Birmingham, B15 2RT
T: 0121 414 6215 F: 0121 414 7051

Lead: James Raftery
Other member: Hugh McLeod

Main Responsibilities: Activity changes in inpatient services; contracting methods; service costs and purchaser efficiency.

HEALTH SERVICES RESEARCH UNIT, LONDON SCHOOL OF HYGIENE AND TROPICAL MEDICINE
Keppel Street, London, WC1E 7HT
T: 0171 927 2231 F: 0171 580 8183

Lead: Colin Sanderson (with Jennifer Dixon, King's Fund)
Other member: Peter Walls

Main Responsibility: A&E services and emergency admissions.

LSE HEALTH. LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE
Houghton Street, London, WC2A 2AE
T: 0171 955 7540 F: 0171 955 6803

Lead: Gwyn Bevan

Main Responsibilities: Resource allocation methods (Bevan); service costs and purchaser efficiency (Le Grand).

ISBN 1-85717-176-4