

Moving towards local allocation of resources

Roy Carr-Hill, Nigel Rice, Peter Smit

Centre for Health Economics - University of York - UK

Introduction

We should first of all recall that one of the overall purposes of formula allocation is to ensure equity of misery whether this is :

- between areas,
- between providers,
- between purchasers.

This is not just an evening cabaret joke but a consequence of the intention to move towards decentralisation : if that is a serious intention then we also intend that the local populations believe that they are equally disfavoured and so will make similar complaints via « democratic » representatives. However, the specific purpose of this paper is to :

- describe some of the changes in England since the 1974 RAWP (introduced in 1976) and the \sqrt{SMR} ,
- examine the implications of local autonomy in the British context.

Finally, we should be cautious about comparisons between England and France.

- the English pragmatism implies that there is no connection between financial planning and auditing (specifically between mechanisms for resource allocation and examining what is done with those resources).

The original 1976 Resource Allocation Working Party (RAWP) enunciated two fundamental principles :

- adjustments should be made for differences in perceived need,
- account needs to be taken of unavoidable geographical differences in costs of providing services.

They decided not to base the adjustment for need on utilisation because of the well known dangers of supplier induced demand. Instead they adopted the following syllogism.

- morbidity measures need
- mortality is closely correlated with morbidity
- mortality is an approximation to need

On this basis, they used the SMRs for fourteen different condition-groups as the basis for calculating allocations. At the time, this was very reasonable syllogism because a substantial fraction of deaths were due, at least in part, to poverty. However, the numbers of deaths in an area now are as much a function of patterns of migration of elderly people.

There were academic criticisms : thus the presumption that there was a direct one-to-one association was seen as inappropriate and the differentiation between condition-specific SMRs unrealistic.

Nevertheless, the formula was used to designate targets for allocating resources to the then existing 14 Regional Health Authorities¹. It was successful in the sense of reducing the distance of each Regional Health Authority from it's target. However, there were many districts within Regions which suffered and specifically districts in and around London began to complain. Moreover, there began

¹ Although not a statutory requirement, it was also used as the basis for allocating from Regions to their constituent Districts.

to be some speculation - only in part ironic - that a formula based only on the number of deaths could have some perverse incentives.

A review was therefore commissioned in 1988, where the proposal was to develop a formula based on the analyses of utilisation data whilst avoiding the problem of supplier-induced demand identified by the RAWP.

However, this review was inadequate for several reasons :

- the episode measure of utilisation used made no allowance for variations in case mix or in severity,
- there was inadequate recognition of different roles played by supply variables,
- the associations between socio-economic characteristics and supply were ignored,
- methodologically and statistically poor.

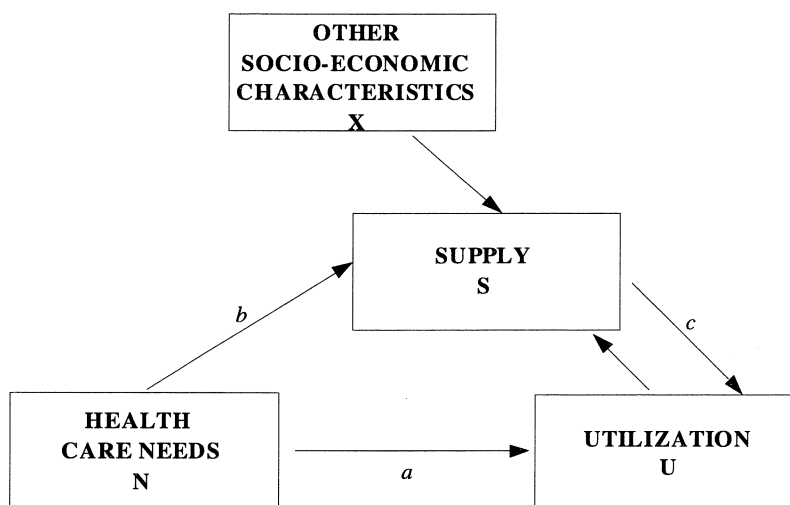
1. Importance of Models

The multiple interaction between need, supply and utilisation have to be properly modelled. And models have to be :

- credible, theoretical accounts of the relationships between need supply and utilisation,
- statistically robust estimates which explain substantial fractions of variations in utilisation.

We take as the general model summarising in an over-simplified form our knowledge about these relationships, the schema outlined in Figure 1.

Figure 1 : The simplified model of demand for care



The problem is to distinguish between S_i and S_{NL} . Algebraically, we can write the following equations. We proceed as follows :

$$U = f(N_R, S_L, S_{NL}, X_1) \quad (1)$$

$$S_L = g_L(U, N_R, X_{2L}) \quad (2)$$

$$S_L = g_{NL}(U, N_I, X_{2NL})$$

Because of the simultaneity in this system of equations, we substitute the estimated values of S_L and S_{NL} to obtain :

$$U = f(N_R, S_L, S_{NL}, X_1) \quad (3)$$

where X_1 summarises variation between districts in policy and practice and

$$\sqrt{V}_L = h_1(N_{R,t-1}, X_{2L}) \quad (4)$$

$$\sqrt{V}_{NL} = h_2(N_{I,t-1}, X_{2NL})$$

If N_I is orthogonal to N_R , then we can rewrite this as:

$$U = f(N_R^+, X^+) \quad (5)$$

where $N_R^+ = \{N_{R,t}, N_{R,t-1}\}$ $X^+ = \{X_1, X_{2L}, N_I, X_{2NL}\}$

Of course, we don't know what are the components of N_R and, especially how to distinguish N_R from N_I .

We therefore proceed in two stages : **first**, we identify the components of N_R by carrying out a classic two stage least squares estimation and **second**, we estimate the coefficients of N_R in a multilevel model in order to control for variations between areas without reintroducing the distorting effect of the supply variables.

This schema needs to be filled out with variables and estimated at an appropriate level. This is where the next set of problems arise.

For the work carried out to develop the « York » formulae for acute and psychiatric in-patient utilisation, the following data were collected at the level of « synthetic electoral ward »² administrative units with average populations of 10,000 and at the district level :

- all admissions (c 9 million) and lengths of stay in 12 speciality groups from the Hospital Episode System,
- costing data for these 12 speciality groups from a study in East Cheshire,
- locations of all hospitals, general practitioners in the country,
- births and deaths in different age groups,
- socio-economic characteristics of the wards from the 1991 Census.

These data were used to construct, for each Synthetic Electoral Ward :

- costs of inpatient utilisation for different speciality groups,
- accessibility scores based on quantity of supply and distance between the SEW and each point of supply for acute beds, non-acute beds, general practitioners, residential homes and private facilities,
- morbidity, mortality and socio-economic indices.

A range of estimation and sensibility checks were carried out with the following general results (for details see Carr-Hill et al 1994).

² Some Electoral Wards were judged too small and combined with neighbouring wards to attain a minimum size of 5,000.

- reasonably well-specified models were estimated in order to identify the 'real' needs variables with specialities grouped into acute and psychiatric³,
- the « needs » variables identified for the acute inpatient utilisation model were :
 - Access to NHS acute beds,
 - Access to general practitioners,
 - Proportion of population aged 75+ not in nursing or residential homes,
 - Access to private hospital beds,
 - Persons divided by hectares,
 - Proportion in households with head in manual social classes,
 - Proportion of pensionable age living alone,
 - Proportion of dependants in single carer households,
 - Proportion of economically active unemployed,
 - Proportion in private rented accommodation,
 - Proportion not in black ethnic groups,
 - SMR for ages 0-74,
 - Standardized illness ratio for ages 0-74.
- the « needs » variables identified for the psychiatric model were :
 - Access to NHS non-acute beds,
 - Access to general practitioners,
 - Proportion of population aged 75+ not in nursing or residential homes,
 - Access to private hospital beds,
 - Proportion in households headed by a lone parent,
 - Proportion of dependants with no carer,
 - Proportion in persons born in New Commonwealth,
 - Proportion of pensionable age living alone,
 - Standardized mortality ratio (SMR) for ages 0-74,
 - Proportion of adult population permanently sick,
 - Percentage of population living in « urban » enumeration districts (as defined by Department of Environment),
 - Proportion of 17 years olds who are students.
- forty per cent of the unexplained variance in utilisation between wards remained at the district level.

These models have been judged sufficiently robust to form the basis for the allocation of resources for acute and psychiatric hospital care from the centre to directly to the 90 purchasing District Health Authorities. Two points are worthy of note in the context of this colloquium :

- in discussing these models with the civil servants responsible for purchasing, ministers and then for implementing the formulae, the economic and sociological assumptions behind the technical choices made were discussed and agreed before exemplifying the models in financial terms.
- although separate formula were estimated for acute and psychiatric care (and later for some other sub-sectors), the financial allocations were NOT made separately thereby devolving the entire responsibility for deciding how to spend health care resources to the District Health Authority.

³ It proved impossible to develop adequate models for other more acute specialities such as geriatric, maternity and mental handicap.

2. The Move to a Primary Care led NHS

There has been considerable rhetoric for a primary care led NHS since the 1990 reforms with the introduction of General Practice fundholders who were allocated sufficient capitation monies to purchase non-urgent hospital treatment in addition to providing primary care services. The new Labour Government, based on what were called Total Purchaser Pilots introduced by the previous government, proposes to extend this so that general practices will be grouped into 500 Primary Care Groups (PCGs) responsible for between 50,000 and 250,000 patients (with an average of 100,000) and allocated an appropriate share of ALL health care resources. The District Health Authority will retain the functions of audit and control, but will only take on the role of purchaser as a last resort, if the Primary Care Groupings are unable to function effectively.

Whilst, in principle, the move towards more local decision-making is to be welcomed⁴, this reforms due to be implemented in April 1989 - poses another set of problems for equitable distribution. Whilst - as above - it is agreed that budgets should capture all expenditure consistent with the move towards Primary Care.

- there are problems with setting equitable budgets,
- some functions, e.g. prevention might be given lower priority as patients demand immediate attention,
- there might be « cream-skimming » of high risk cases by the PCG in order to avoid excessive expenditure,
- there is likely to be unpredictable year-on-year variations,
- lack of information,
- management cost may be high.

And these problems are additional to the difficulties of controlling for the impact of supply on utilisation.

The problem of safeguarding prevention can be avoided by allocating specific amounts for those activities; the solution currently adopted with GP fundholders of supplementing allocations for cases with expenditure above quite a low ceiling appears to have avoided the problems of cream-skimming (Glennerster, 1998) ; and the issue of high management costs is not a resource allocation problem. The remainder of this paper is therefore concerned with the problem of setting equitable budgets in the presence of « unpredictable » year on year variations.

3. Problems of Setting Equitable Budgets for Primary Care Groups in England

3.1. List Inflation

Patients in England are registered with a single General Practitioner and these registrations are compiled into a single data base in order to work out, capitation payments (which form 60% of incomes for GPs and 100% for GP fundholders) and the Jarman deprivation payment given to GPs in respect of patients living in deprived areas⁵.

⁴ *Not only is small beautiful, it is also sensible. The only technical reason, in addition to those considered below about unpredictable variations one would not advocate each individual holding at least part of their health care budget is that they will not in general have sufficient information.*

⁵ *The Jarman index is the weighted sum of eight Census variables, standardised and transformed with an arc sine transformation, the variables having been identified and the weights chosen through a national survey of General Practitioners. There is no obligation of the GP to spend the extra money on those patients.*

A long recognised curiosity - which now has to be resolved - is that there are 53 million people registered with GPs but the current population of England based on the last Census is only 50 million. This would not be very important if it were uniform across England because then a uniform scaling factor could be applied. However, list inflation ranges from -4% in Cambridge and Huntingdon (the constituency of our previous Prime Minister John Major) to +30% in Camden and Islington (the constituency of our current Minister of Health, Frank Dobson). Recent attempts to identify practices which are slow at communicating a death or a transfer of registration have failed and so these are potentially massive movements in budgets as allocations have to be set on the basis of lists. The likely resolution is that there will be agreed scaling factors applied to patients living in different areas so that the total scaled population will still be 50 million.

3.2. Attribution

In order to develop formulae, we need to identify indicators which are :

- useful - an acknowledged risk factor
- reliable - not subject to major errors
- universal - available for all patients
- objective - not subject to substantial variations in judgement
- up-to-date - capturing contemporary characteristics of patients
- free of perverse incentives.

The problem in England, is that there is very little data kept - routinely about the individual (partly because we do not have n identify card) essentially limited to age, sex and address. The procedures adopted in deciding whether or not a GP is currently entitled to a Jarman deprivation payment (see above) is therefore to « attribute » the characteristics of the enumeration district or electoral ward in which the patient lives to the patient. This has the advantage of making it possible to use much more data to the individual (such as the socio economic characteristics and mortality data used in the analyses of hospital in-patient utilisation) but has the potential disadvantage that the individuals from a specific small area on a particular GPs list may not be typical of the area in which they live (the ecological fallacy). Unfortunately, analyses have shown that this is the case whilst it is not easy to see why this should happen (because, on the whole, individuals register with their nearest GP rather than according to any other more sociological criteria) analyses by, for example, Carr-Hill and Rice (1995) have shown that there is need a substantial mismatch between individuals and areas (see Table 1). Work is currently being carried out to find the most appropriate methodology for minimising these biases.

3.3. Unpredictable Variation

There are five possible sources of variations on expenditure :

- variation predicted by the formulae,
- variation predictable using some set of needs indicators but not captured by the formulae,
- variation due to clinical practice of GPs,
- variations due to local health care policy,
- variations which is truly random,

Substantial variations in expenditures from the allocated budget could have serious consequences such as :

- low spenders « spending up » to their budget on unnecessary items,

- high spenders might constrain treatment options for their patients and then react with hostility against formula allocation,
- patients with identical needs might be treated differently,
- GPs might join 3rd party insurance schemes with no value added for the health service.

Clearly some risk pooling will be essential but some of those problem items from the inadequacy of a formula based on area data.

4. Towards a European Wide Health Care Utilisation Survey for Local Planning

The only way of resolving some of the debates about the adequacy of formulae that might be proposed is to carry out a longitudinal study of health care utilisation related to prior health characteristics. If we also assume that.

References

- 1 Carr-Hill, R., Hardman, G., Martin, S., Peacock, S., Sheldon, T. and Smith, P. (1994). A Formula for Allocating Health Care Resources Based on a Small Area Analysis of Hospital In-Patient Utilisation, Centre for Health Economics, University of York (Occasional Paper)
- 2 Carr-Hill, R. and Rice, N. (1995) « Is the Enumeration District or Ward Most Appropriate for Attribution » Journal of Epidemiology and Community Health.
- 3 Glennester, H. (1998) « Health Care Policy and Practice » being published in special issue of International Journal of Quality in Health Care.
- 4 Jarman, B. (1983). « The Under-Privileged Area Score » British Medical Journal