synthesised, then we are in the realms of meta-analysis. An amusing account of the pitfalls in the analysis was given by Huitema.[6]

In conclusion, n of 1 trials are best applied when there is genuine doubt about the best treatment for a stable condition in individual patients and when the treatments have already been well investigated in clinical trials. They are not a substitute for a properly controlled phase III trial to decide the efficacy of a new treatment.

1 Johannessen J. Controlled trials in single subjects. I. Value in clinical medicine. *BMJ* 1991;303:173-4.
2 Lewis JA. Controlled trials in single subjects. II. Limitations of use. *BMJ* 1991;303:175-6.
3 Senn SJ. Suspended judgement: n-of-1 trials. *Control Clin Trials* 1993;14:1-5.
4 Sunderland A. Single case experiments in neurological rehabilitation. *Clinical Rehabilitation* 1990;4:181-92.
5 Crabtree BF, Ray SC, Schmidt PM, O'Connor PJ, Schmidt DD. The individual over time: time series applications in health care research. *J Clin Epidemiol* 1990;43:241-60.
6 Huitema BE. Autocorrelation in behavioral research: wherefore art thou? and statistical analysis and single subject designs: some misunderstandings. In: Poling A, Fuqua RW, eds. *Research methods in applied behavior analysis.* New York: Plenum Press, 1986:187-232.

Allocating resources to health authorities: development of method for small area analysis of use of inpatient services

Roy A Carr-Hill, Trevor A Sheldon, Peter Smith, Stephen Martin, Stuart Peacock, Geoff Hardman

Every year about £22 billion is allocated to health authorities for hospital and community services in England. The distribution of most of these funds is based on a formula developed to reflect the population's needs, but the existing formula has been criticised on several grounds. This paper describes the development of a method to determine the health needs for small geographical areas. Data from the hospital episodes statistics and 1991 census together with information on vital statistics and supply of health care facilities were used in the model. Two stage least squares regression was used to identify true indicators of need, and these were entered into a multilevel model to take account of variations in practice in different health authorities. The resulting formula should be more statistically robust and more sensitive to needs than previous approaches.

The long debate in the United Kingdom about how NHS funds should be distributed was reviewed by Mays and Bevan.[1] In the hospital and community sector allocations have been made on the basis of formulas. In the 1970s the Resource Allocation Working Party (RAWP) was established to recommend a system for allocating resources according to the health needs of the population and to identify and correct inequalities in distribution.[2] It recommended distributing funds on the basis of population, weighted according to differences in the need for health care and unavoidable geographical differences in the costs of providing services. The principle of a weighted capitation formula has remained intact since.

Because considerable variations exist in the use of NHS resources by different age and sex groups, the population of each area was separated into age and sex groups. These groups were weighted by an index of morbidity that reflected differences in health care needs caused by factors apart from demography. As no better indices existed the working party recommended using disease specific standardised mortality ratios as a proxy for relative needs.

The recommendations were used as the basis for allocations to the 14 English regional health authorities until 1990. The allocations determined by the Resource Allocation Working Party methods were initially used as targets, to which actual allocations were to converge over several years. Many regions used its principles as the basis for distributing money to district health authorities.

In 1985 the government set up a review of the Resource Allocation Working Party system to improve the accuracy with which the formula measured relative need.[3,4] Most of the work was based on a stepwise ordinary least squares regression analysis of the determinants of hospital use in small areas (electoral wards) after statistical methods had been used to adjust for variations in the supply of acute beds.[5] The review recommended the use of the all cause standardised mortality ratio for the under 75 age group and including an index of social deprivation.

The government implemented only part of the review's recommendations. The resulting regional formula, which remains in use, is based on the square root of all cause standardised mortality ratio for those under 75 as an index of relative need but does not contain an index of social deprivation. Most regions have adopted a similar approach for distributing to districts, although the formulas vary from region to region.

The purpose of this study was to develop a more sensitive empirically based model of demand for hospital inpatient care for small areas. The study formed part of the Department of Health's review of the weighted capitation formula used to distribute funds for hospital and community services to regional health authorities in England. We used the underlying principles of previous work but tried to overcome some of the statistical and other shortcoming identified in the review.[6,7] This paper describes the methods used and the results and implications are described in an accompanying paper.[8]

Theoretical framework

Ideally, an empirical study of relative need in small areas would be based on morbidity data. But the only source of such data in Britain—the general household survey—collects only limited information and has a relatively small sample size. Therefore any practicable empirical study has to adopt the major assumption that some measure of use of health care can be used to predict health care needs.[9] This presumes that the legitimate health care needs of all care groups are reflected equally in measures of use. For this to be true the existing pattern of allocations to client groups must be satisfactory, and the chosen measure must reflect those allocations. In practice, neither of these requirements is met[10] but the data currently available are insufficient to use alternative methods. The methods that we used to attempt to solve part of this problem are discussed later.

We considered two types of determinant of demand to be important: the health needs of the population and the supply of health care facilities. Needs were broadly interpreted to include indirect social determinants of demand for health care as well as direct measures of health. Supply was included since it is widely believed that the availability of health care services affects

Centre for Health Economics, University of York, York YO1 5DD
Roy A Carr-Hill, *senior research fellow*
Trevor A Sheldon, *senior research fellow*
Geoff Hardman, *research fellow*

Department of Economics and Related Studies, University of York, York YO1 5DD
Peter Smith, *senior lecturer*

Institute for Research in the Social Sciences, University of York, York YO1 5DD
Stephen Martin, *research fellow*

York Health Economics Consortium, University of York, York YO1 5DD
Stuart Peacock, *research fellow*

Correspondence to:
Mr Smith.

BMJ 1994;309:1046-9

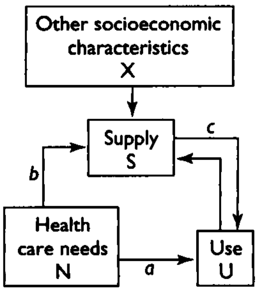

FIG 1—*Simplified model of demand for health care*

demand in two ways. Firstly, when there is excess demand, supply constraints affect the care that can be offered, and, secondly, supply of doctors can induce demand.[11] As well as influencing use, supply may itself be influenced by past use and needs.[6,7] Figure 1 shows how health care needs might directly influence use of health care, which is also mediated partly by supply. In addition, some of the variation in supply is determined by the distribution of health care needs, other socioeconomic conditions, and past use, either because funding has been based on these variables or because of other historical relations that determine supply. An important aim of this study was therefore to separate the effects of supply and needs on use.

If supply affects current use but is also influenced by past use and need it is inappropriate to use conventional ordinary least squares regression methods to model the determinants of use of health services. Instead, a method such as two stage least squares is needed to take account of the simultaneous determination of use and supply (appendix 1).[12]

Collection of data

The units of analysis used in the study were 4985 "synthetic wards," small areas with average populations of 9643 covering the whole of England. These synthetic wards were electoral wards, aggregated with contiguous wards where necessary so that none had a population of less than 5000. For each synthetic ward data were assembled relating to socioeconomic conditions, the supply of health services, and the use of inpatient services.

The socioeconomic variables comprised detailed demographic data prepared by the Office of Population Censuses and Surveys, health status variables, and broader social and economic variables derived from the 1991 census. The demographic data were used to standardise all variables for which age was thought to be an important determinant, as follows. The national rates M_{jk} for the variable for age group j and sex k were applied to local population sizes p_{ijk}. This gave the equation

$$EN_i = \sum_{j=1}^{18} \sum_{k=1}^{2} M_{jk} p_{ijk}.$$

when EN is the expected number of observations of the phenomenon of interest in area i. The indirectly standardised rate is then the ratio of the actual number of observations to this expected number. We found little difference between values and rankings of wards using a direct standardisation method and indirect standardisation.

The health status variables included a variety of age specific standardised mortality ratios, standardised ratios of self reported limiting longstanding illness (derived from the 1991 census), and low birthweight data. We abstracted data from the census on 37 socioeconomic variables that could possibly influence demand for health care. In summary, the census variables covered the following aspects of social and economic circumstances: housing tenure; housing amenities; car ownership; overcrowding; ethnicity; country of origin; elderly living alone; lone parents; students; migrants; unemployment; educational qualifications; social class; and non-earning households. Clearly several measures could be added to these. But we believe that for the purposes of this study the range of issues covered is sufficient to capture the important social causes of the need for health care.

We created four supply variables to capture the effect of varying availability of health services to the wards' populations. These sought to measure the accessibility of NHS inpatient facilities, general practitioner services, the provision of residential and nursing homes, and the accessibility of private inpatient facilities. When deriving measures of accessibility it is necessary simultaneously to reconcile the supply of facilities, their proximity to the ward of interest, and the impact of competing populations and competing supply. This was done by using the methods of spatial interaction modelling (appendix 2).[13]

To develop measures of accessibility, we used the following indicators of the size of service provision: the number of available beds for NHS inpatient facilities; the number of general practitioners for general practitioner services; the number of patients present on census night for private hospitals. Hospitals and general practitioner surgeries were located using grid reference, and we calculated distances to populations on the basis of straight line distances to ward centroids. The proportion of those aged 75 or over living in nursing or residential homes was the remaining health care supply variable. Limitations of data and the time scale of the project precluded development of more refined variables.

Rates of use standardised for age and sex were calculated from the hospital episode statistics, which cover all finished inpatient and day case hospital episodes. We used 8 566 887 valid records for the financial year 1990-1 and 9 042 169 records for 1991-2. The data made available to us contained the following information: district of treatment; method of admission; source of admission; category of patient (public/private); wait for elective admission; age group; speciality group; operation group (×4); order number of episode; duration of episode; discharge destination; patient classification (day/ordinary); synthetic ward of residence.

Ideally the analysis would have taken into account the variability of use of hospital care by patients with specific diagnoses or for specific procedures. However, all diagnostic and procedure data were deleted from the hospital episodes statistics before they were sent to us. Instead we assigned each episode to one of 12 specialty groups.

We had estimates of fixed costs per episode and variable costs per bed day for the 12 specialty groups prepared by East Cheshire Statistical Analysis Consultancy. We used these to attach a cost to each episode and measure a ward's use of service in terms of costs. Two types of cost were calculated for each episode: the standard cost is the national average cost for a particular age, sex, and specialty group; the estimated cost is the specialty-specific cost for the length of stay of the episode—that is, speciality fixed cost + speciality variable cost × length of stay. Standard costs seek to remove local variations in policy and practice from the measure of use, but do not capture variations in lengths of stay brought about by variations in needs. For this reason, the Department of Health technical advisory groups recommended using estimated costs as the basis for the analysis of acute episodes. The table shows the fixed and variable costs for each of the 12 specialty

Specialty costs derived from East Cheshire Statistical Analysis Consultancy

Specialty group	Description	Cost per episode (£)	Cost per bed day (£)
Surgery	All surgery excluding neurosurgery, plastic surgery, cardiothoracic surgery, and paediatric surgery	153.30	165.10
Surgery II	Neurosurgery, plastic surgery, cardiothroracic surgery, and paediatric surgery	253.80	277.10
Medicine	All medical excluding geriatrics, cardiology, medical oncology, neurology	149.50	123.60
Geriatrics	Geriatric medicine	0.00	98.80
Medicine II	Cardiology, medical oncology, neurology	560.30	111.80
Psychiatric	Psychiatric	1031.30	86.50
Learning disabilities	Learning disabilities	1031.30	85.90
Maternity	Maternity	144.10	147.30
Gynaecology	Gynaecology	144.10	153.70
Radiotherapy	Radiotherapy and radiology	144.10	159.00
Other	All other valid episodes with code	144.10	91.80
Not stated	All episodes with invalid code	181.20	119.80

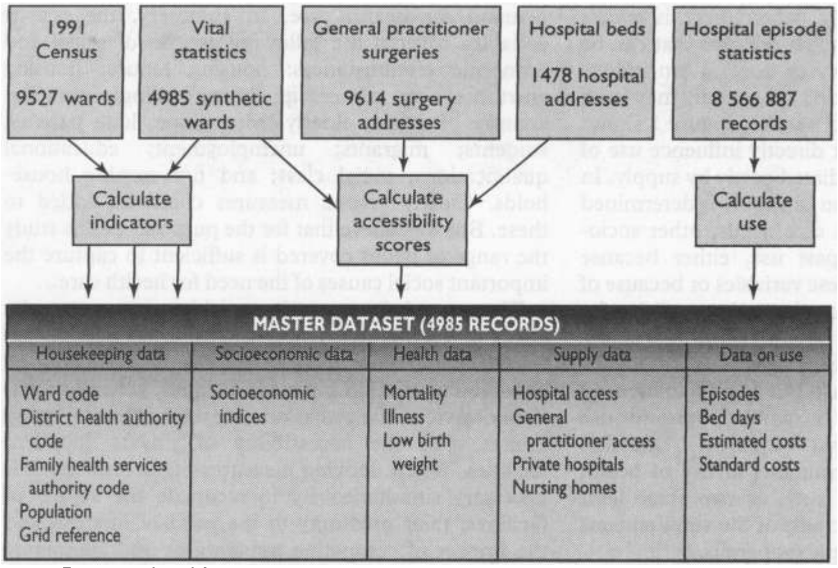

FIG 2—*Representation of the construction of master dataset for model of use of health services*

groups used in the study. The cost data also facilitated the construction of age-cost curves for inpatients, defined as the national average costs per head of population within age bands. Figure 2 shows the links between the various data sources and the master database.

Adjustment for errors

Inevitably, such a large dataset constructed from a variety of independent sources will contain errors. Differences exist between districts over what constitutes a finished consultant episode.[14] However, the degree to which this might result in systematic differences between wards is unknown. In addition, a comparison of the hospital episode statistics for each district with the statistical return KP70, which provides independent estimates of total annual episodes, showed discrepancies in some regions (more in 1990-1 than in 1991-2). We took KP70 data as the standard, and when constructing the measures of use we multiplied each episode by a factor defined as the ratio of the number of completed episodes from the KP70 return to that derived from the hospital episode statistics for the district in which the episode took place.[15]

The one systematic and potentially significant coding error in the hospital episodes statistics that we examined concerned the use of a dummy postcode when the correct postcode was not known. In certain districts clerks have allocated all inpatients with unknown postcodes to a "dump" postcode within the district. This resulted in a large number of episodes being assigned to the ward which contained the dump postcode. Examination of heavily used postcodes suggested wards with large scale dumping, and these were omitted from the analysis. When dumping occurred there was likely to be a concomitant under-recording of use in neighbouring wards. We could not adjust for this phenomenon except to delete a small number of wards with exceptionally low rates of use. A total of 45 of the 4985 wards were omitted from the analysis.

Modelling methods

Initially, we carried out tests to determine whether there was simultaneity between supply and use (a test for endogeneity).[16] Since endogeneity was found in all models, we modelled use of health care for the wards as a function of supply and needs by two stage least squares regression (appendix 1). Numerous separate models were estimated for acute, geriatrics, psychiatry, learning disabilities, and maternity specialty groups. The models performed better when natural logarithms were taken of all variables. Each observation was weighted in proportion to the total population of the ward to ensure that the national average model of resource use did not give undue weight to patterns of use in smaller synthetic wards. All measures of health and socioeconomic variables were included as instruments in the modelling.

In addition to the supply variables, we also included several potential determinants of health care needs in the model. Needs variables were initially selected on the basis of statistical significance and prior judgment about the importance of the variable. Thus, for example, in the acute sector a model comprising four supply variables and 25 needs variables was specified as the unrestricted model. This model was then progressively restricted by omitting needs variables that were not significant or that had small standardised effects on predictions (β coefficients).[17] This process was continued until deletion of another variable would have altered the model significantly. Tests were made to ensure that the model was statistically well specified.[18]

Developing a resource allocation formula

To measure underlying need for health care the analysis had to allow for the effect on use of variations in supply. Effectively this means assuming that all supply in an area is at some national average level appropriate to the needs of the population. The variation in use due to variation in supply variables should therefore be incorporated only to the extent that supply reflects variations in legitimate need for health care. We needed to develop a measure of "normative use" (the use in an area if the response to its needs was at the national average level).

Consider the model of use illustrated in figure 1. Variations in use (U) arise because of variations in needs (N) and variations in supply (S). Normative use is that part which is attributable to needs alone. Needs can influence use in two ways: firstly, directly, as indicated by the arrow *a*, and, secondly, through supply, since some decisions on where to invest in supply in the past were influenced by needs (arrows *b* and *c*). The analytical task is to find that part of the supply effect *c* which is attributable to factors (X) other than needs and to remove that part from the model.

The two stage least squares modelling exercise isolated legitimate needs that had an unambiguous statistical relation with use. However, the coefficients on the supply variables will reflect both legitimate needs (N) and extraneous variables (X). To isolate the impact of needs on use, both directly and as mediated through supply, the legitimate needs variables (identified from the two stage regression analysis) were therefore entered in an ordinary least squares regression of the cost of use as follows:

$$U_i = \alpha + \sum_{j=1}^{m} \beta_j N_{ij} + \varepsilon_i.$$

The coefficients in this equation reflect the total impact of needs on use, and can be used as the basis for a resource allocation formula.

This approach is different from that adopted in the review of the Resource Allocation Working Party system, which simply took the regression including both needs and supply variables and ignored the supply variables.[5] Our strategy allows for the fact that variations in existing supply might already to some extent reflect variations in legitimate health care needs.

A final consideration was that there are likely to be systematic effects of administrative areas (such as health authorities) on rates of use. This hierarchical structure to the data contravenes a fundamental assumption of conventional statistical modelling: that the unexplained variation is independently distributed. Such clustering might lead to incorrect inferences when using ordinary regression methods. Therefore, we used multilevel modelling techniques to take account of hierarchical clustering of the errors and to try to distinguish between interdistrict and intradistrict variations.[19 20] More details of the methodology used in this study are reported elsewhere.[21]

This work was funded by the Department of Health as part of its review of weighted capitation in England. We thank everyone who helped. In particular, the study team thanks the members of the technical group and steering group, Professor Harvey Goldstein and his team, Institute of Education; Chris Orme, University of York; Michael Borowitz, Battelle Europe; and George Davey Smith, University of Glasgow.

1 Mays N, Bevan G. *Resource allocation in the health service.* London: Bedford Square Press, 1987. (Occasional paper in social administration No 81.)
2 Department of Health and Social Security. *Sharing resources for health in England: report of the Resource Allocation Working Party.* London: HMSO, 1976.
3 Department of Health and Social Security. *Review of the Resource Allocation Working Party formula: interim report by the NHS Management Board.* London: DHSS, 1986.
4 Department of Health and Social Security. *Review of the Resource Allocation Working Party formula: final report by the NHS Management Board.* London: DHSS, 1988.
5 Coopers and Lybrand. *Integrated analysis for the review of RAWP.* London: Coopers and Lybrand, 1988.
6 Mays N. NHS resource allocation after the 1989 white paper: a critique of the research for the RAWP review. Community Medicine 1989;11:173-86.
7 Sheldon TA, Carr-Hill RA. Resource allocation by regression in the NHS: a statistical critique of the RAWP review. *Journal of the Royal Statistical Society [Am]* 1992;155:403-20.
8 Smith P, Sheldon TA, Carr-Hill RA,Martin S, Peacock S, HHardman G. Allocating resources to health authorities: results and policy implications of small area analysis of use of inpatient services. *BMJ* 1994;309:1050-4.
9 Sheldon TA, Davey Smith G, Bevan G. Weighting in the dark: resource allocation in the new NHS. *BMJ* 1993;306:835-9.
10 Morgan M, Mays N, Holland W. Can hospital use be a measure of need for health care? *J Epidemiol Community Health* 1987;41:269-74.
11 Cromwell J, Mitchell JB. Physician-induced demand for surgery. *Journal of Health Economics* 1986;5:293-313.
12 Hancock KE, Holden DR, Swales JK. Consistency of the spatial allocation of NHS resources: an econometric analysis. *Applied Economics* 1991;23: 1623-36.
13 Wilson AG. *Urban and regional models in geography and planning.* Chichester: Wiley, 1974.
14 Clarke A, McKee M. The consultant episode: an unhelpful measure. *BMJ* 1992;305:1307-8.
15 Department of Health. *Hospital episode statistics: vol 1 England 1989/90.* London: DoH, 1993.
16 Maddala GS. *Econometrics.* Singapore: McGraw-Hill, 1987.
17 Hendry DF. Econometric methodology: a personal perspective. In: Bearly TF, ed. *Advances in econometrics: fifth world congress.* Cambridge: Cambridge University Press, 1987.
18 Godfrey LG. *Misspecification tests in econometrics.* Cambridge: Cambridge University Press, 1988.
19 Paterson L, Goldstein H. New statistical methods for analysing social structures: an introduction to multilevel models. *British Education Research Journal* 1991;17:387-93.
20 Gatsonis C, Sharon-Lise N, Lin C, Morris C. Geographic variation of procedure utilization. *Medical Care* 1993;31:YS54-9.
21 Carr-Hill RA, Hardman G, Martin S, Peacock S, Sheldon TA, Smith P. *A formula for distributing NHS revenues based on small area use of hospital beds.* York: University of York, 1994. (Centre for Health Economics occasional paper.)

(Accepted 22 March 1994)

Appendix 1

TWO STAGE LEAST SQUARE MODEL

The central hypothesis in this study is that use of health services depends on needs and health care supply. Health care supply in turn depends on needs and use. Therefore a simultaneous equations model is suggested—a standard econometric technique used when endogeneity is present. It entails first regressing each of the supply variables (S) on the set of instruments (variables) and then using the predicted values of the supply variables as explanatory variables in the regression in place of their actual values. This approach should give consistent estimates of all coefficients in the regression of use on needs and supply. Standard errors are adjusted to take account of the endogeneity.

In practice, we found that the supply variables were indeed endogenous and that a method such as two stage least squares was needed. Although the instrumented supply variables are used in estimating all the coefficients in the equation, the coefficients so estimated apply to the original supply variables and not to the instrumented supply variables.

If use (U_i) in area i is a function of needs (N_i) and supply (S_i) in area i then $U_i=f(N_i, S_i)$. However, if supply is influenced by use, needs, and other determinants (X_i) then $S_i=g(U_i, N_i, X_i)$. Eliminating supply S from these two equations gives the following equation: $U_i=g_i(N_i, g_2(N_i, U_i, X_i))$. This can be solved to give the following expression for use: $U_i=g_3(N_i, X_i)$. This last equation is known as the reduced form expression for U, which explains U in terms of legitimate needs variables N and more general socioeconomic variables X. The impact of needs (N) on use (U) is given by the total derivative:

$$\frac{dU}{dN}=\frac{\delta g_3}{\delta N}+\frac{\delta g_3}{\delta X}\frac{\delta X}{\delta N}$$

That is, the total effect of needs on use is found by examining both the direct effect of N on U, and any indirect effect on U associated with X, if X is correlated with N.

MULTILEVEL MODELLING

The basic structure of the multilevel model is the same as the conventional statistical model. However the regression equation taking into account legitimate needs

$$U_i=\alpha+\sum_{j=1}^{m}\beta_j N_j+\varepsilon_i$$

is extended to the following:

$$U_{ik}=\alpha+\sum_{j=1}^{m}\beta_j N_{ijk}+\varepsilon_{ik}+\upsilon_k$$

where k represents a district (or other administrative area). Instead of assuming that a single value exists for each of the β and γ parameters to be estimated, it is assumed that the parameters might vary between administrative areas. In effect, for each explanatory variable in the model the multilevel analysis produces an estimate of the slope and intercept specific to each administrative area.

Appendix 2

METHOD OF CONSTRUCTING SUPPLY VARIABLES

For NHS inpatient facilities the accessibility A_i of zone i to hospital facilities is given by the complex expression

$$A_i=g\left(\sum_d T_{id}\right)/P_i=g\sum_d B_d S_d f(C_{id})=g\sum_d\left(\frac{S_d f(C_{id})}{\sum_r P_r f(C_{rd})}\right)$$

where T_{id} is the number of interactions (hospital episodes per year) between residential zone i and hospital d; P_i is the population of zone i; S_d is some measure of the size or attractiveness of hospital d; B_d is a balancing factor for hospital d; c_{id} is some measure of distance (or time) between i and d; f(.) is a distance decay or deterrence function; and g is a constant.

The equation, which is similar to the supply measure used in the review of the Resource Allocation Working Party system can be interpreted simply as the ratio of population (weighted by distance) to hospital size (weighted by distance). If the measure of hospital size is taken to be beds the equation is directly analogous to the familiar beds per head ratio but takes account of distance and competition from other wards.

In the hospital episodes statistics provided by the Department of Health the hospital of treatment was deleted and only district of treatment given. Thus more sophisticated catchment area analysis could not be carried out.