This is a repository copy of Risk adjustment for hospital use using social security data: cross sectional small area analysis.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/132/

**Article:**

https://doi.org/10.1136/bmj.324.7334.390

**Reuse**
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
Risk adjustment for hospital use using social security data: cross sectional small area analysis

Roy A Carr-Hill, James Q Jamison, Dermot O'Reilly, Michael R Stevenson, James Reid, Barry Merriman

Abstract

Objectives To identify demographic and socioeconomic determinants of need for acute hospital treatment at small area level. To establish whether there is a relation between poverty and use of inpatient services. To devise a risk adjustment formula for distributing public funds for hospital services using, as far as possible, variables that can be updated between censuses.

Design Cross sectional analysis. Spatial interactive modelling was used to quantify the proximity of the population to health service facilities. Two stage weighted least squares regression was used to model use against supply of hospital and community services and a wide range of potential needs drivers including health, socioeconomic census variables, uptake of income support and family credit, and religious denomination.

Setting Northern Ireland.

Main outcome measure Intensity of use of inpatient services.

Results After endogeneity of supply and use was taken into account, a statistical model was produced that predicted use based on five variables: income support, family credit, elderly people living alone, all ages standardised mortality ratio, and low birth weight. The main effect of the formula produced is to move resources from urban to rural areas.

Conclusions This work has produced a population risk adjustment formula for acute hospital treatment in which four of the five variables can be updated annually rather than relying on census derived data. Inclusion of the social security data makes a substantial difference to the model and to the results produced by the formula.

Introduction

The 1990s saw an increase in managed care in the United States and western Europe. This change was partly in response to growing awareness of the inescapable scarcity of healthcare resources in almost all countries in the Organisation for Economic Co-operation and Development. Various market style approaches to reforming health care have also been tried to help contain costs. At the same time many countries have been trying to improve funding mecha-

isms so that the whole population has equal access to care—for example, Canada, Germany, the Netherlands, the United Kingdom, and the United States. Methods for adjusting funding according to need (risk adjustment) have probably been most carefully studied in the United Kingdom. Equity of funding has been a recurring preoccupation of NHS policymakers and analysts for at least 20 years. There has been periodic and sometimes acrimonious debate about how best to use available morbidity and socioeconomic data to reflect health care needs once demographic differences have been accounted for.

The original English Resource Allocation Working Party report in 1974 recommended using standardised mortality ratio as a default proxy for morbidity and ultimately need for health care. Definitive empirical analysis of the relation between need and use was impossible at that time because of the lack of comprehensive data on use of health services that were linked to area of residence and because of the systematic confounding of supply of, and demand for, health services. During the 1980s, however, it became increasingly recognised that any risk adjustment formula should include measures of social deprivation as well as health and that the effects of supply of facilities needed to be disentangled from their use so that the relative effects of social deprivation and morbidity could properly be estimated.

Availability of data across the United Kingdom has improved greatly in recent years, and methods to adjust for the confounding of need and supply have been developed. However, previous methods have relied on census data, which are often out of date and include only proxy measures of household income such as car ownership. We describe a study of the determinants of use of inpatient services undertaken as part of a review of the expenditure needs of the four health and social services boards in Northern Ireland. As part of this study we investigated the potential contribution of social security data as direct measures of poverty.

Methods

We assembled large quantities of data on broad population healthcare needs (both health and socioeconomic); use of inpatient services; and supply of hospital...
and community services. We aggregated data on needs and use to electoral ward level (average population 3200) and attached grid references to the supply variables for use in the spatial interactive modelling (see below). When electoral wards were small, we amalgamated neighbouring electoral wards to ensure a minimum population size of 2000.

**Needs**

The health variables included mortality (in the form of standardised mortality ratios), limiting long standing illness and permanent sickness (from the 1991 census), and low birth weight (<2500 g) from the boards’ child health systems for July 1990 to June 1996. There were 34 socioeconomic needs variables, which were mainly drawn from the census. These included religious denomination, which is recognised to be an important social indicator in Northern Ireland. We also included ward data from the end of 1996 on recipients of income support and family credit. (Family credit was paid to families in which the head of household was in a low paid job and has been superseded by the working families’ tax credit.) Recipients of income support were divided into two broad age groups: 18-64 years and ≥65.

**Use of services**

We used routinely available hospital data for 1994-5 and 1995-6 to derive numbers of discharges and bed days for inpatients and day cases by specialty. Non-residents and private patients were excluded. We estimated specialty costs, consisting of a fixed and daily variable component, by regression using data from all hospitals in Northern Ireland. The use and specialty cost data were used to produce a measure of intensity of use at ward level (estimated cost divided by expected cost). The separate funding received by teaching hospitals was discounted.

We adjusted for the size and the age and sex distribution of the population within each ward by indirect standardisation using the overall Northern Ireland rates. For most variables, we used 18 age groups and two sex groups.

**Supply of health services**

We used spatial interactive modelling methods to reflect the influence of supply on usage. These provide a means of reconciling the proximity of each ward to all possible facilities and the attractiveness (usually size) of each facility. We developed distinct models for acute beds (by specialty grouping), private beds in health service hospitals, geriatric beds, care homes, and general practices (including data on the availability and location of branch surgeries and the whole time equivalent number of doctors in a practice). We estimated travel times to hospital using digitised road network data and used these to calibrate the acute specialty models.

**Modelling methods**

Because of the high degree of intercorrelation among the needs variables, we used correlation, cluster, and regression analysis to aid data reduction. We log transformed the needs and supply variables and cost weighted utilisation data to correct for skewness, allowing the use of fully multiplicative regression models.

In modelling hospital use, interactions occur between supply, use, and socioeconomic factors (including lagged interrelations and feedback loops), and this makes it difficult to obtain unbiased estimates of the coefficients for the relation between need and use. We concentrated on disentangling the feedback loop caused by simultaneous supply of, and demand for, health care (endogeneity). This arises because although the physical supply of beds at ward level is responsive to historical demand, historical supply itself may have stimulated use and could also be influenced by factors such as the characteristics of the local area and the general practitioners working within it.

The problem therefore is how to distinguish between the “appropriate” level of supply and extra supply or undersupply. Although a theoretically pure distinction can be made, problems arise in dealing with real empirical data. We argue that as the factors influencing extra supply (deviations from the appropriate level of supply) are at most only weakly correlated with the needs drivers for appropriate supply, the true needs drivers can be identified. However, because of the interrelations between supply and use, the variables have to be identified by purging the estimation of the intercorrelated errors. This is achieved by using two stage least squares (rather than ordinary least squares) regression.

Tests for simultaneity between supply and use (indicating endogeneity) on our data were significant and so we modelled use of hospital services as a function of supply and need by two stage least squares. We then excluded those needs drivers that were found to affect use only through supply, along with the supply variables themselves. The second stage of the regression was concerned with estimating coefficients for the surviving drivers, which were taken to directly affect use. This provided an adjustment for the influence of supply on use.

Some variables are specified only at a higher level than electoral ward (health and social services board). As there are only four boards in Northern Ireland, we used dummy variables in the single level regression.
equations to control for this rather than multilevel modelling.20-21

Table 1 lists the health and social needs variables entered into the regression models as both explanatory and instrumental variables.22 The overall set of variables was reduced until the test for heteroscedasticity was no longer significant. To establish what difference having the social security variables available made, we reworked the modelling without them.

Results

Table 2 gives the full model with both supply and needs variables, although the board dummy variables are not shown. The supply variables that were significantly associated with use of inpatient services were access to hospital beds, general practices, residential and nursing homes, and geriatric beds and use of private beds. The two stage least squares equation for all specialties was significantly endogenous ($F_{ skeptic } = 2.28; P < 0.05$).

A parsimonious model (with five variables) retained most of the explanatory power of the full model (adjusted $R^2 = 52\%$, table 3). This risk adjustment model has been adopted for use in conjunction with an age-sex cost curve for acute hospital services in Northern Ireland to distribute funds for acute hospital services to the health and social services boards. The formula comprises two income related variables, two health variables, and a "social fabric" variable (over 75s living alone). All of these seem intuitively appropriate. The low income indicators supplanted all other socioeconomic indicators.

Table 4 shows the model obtained when the income support and family credit variables were excluded from the candidate set. There was no endogeneity so the modelling was by weighted least squares regression. This model contains seven variables, none of which is related to poverty, although many of the census based socioeconomic indicators are surrogate measures of income and material disadvantage.

Table 5 shows the results of applying the two models to a notional sum of £500m, which is roughly the amount spent on acute services in Northern Ireland annually. The allocations produced using the crude and effective (age weighted) populations are also shown for comparative purposes. Because the size of a population has by far the greatest influence on its need for health care size, any formula of this kind will have an age­sex cost curve for acute hospital services in Northern Ireland to distribute funds for acute hospital services to the health and social services boards. The model components include access to general practice, hospital beds, general practices, residential and nursing homes, and geriatric beds and use of private beds. The supply variables that were significantly associated with use of inpatient services were access to hospital beds, general practices, residential and nursing homes, and geriatric beds and use of private beds. The two stage least squares equation for all specialties was significantly endogenous ($F_{ skeptic } = 2.28; P < 0.05$).

A parsimonious model (with five variables) retained most of the explanatory power of the full model (adjusted $R^2 = 52\%$, table 3). This risk adjustment model has been adopted for use in conjunction with an age-sex cost curve for acute hospital services in Northern Ireland to distribute funds for acute hospital services to the health and social services boards. The formula comprises two income related variables, two health variables, and a "social fabric" variable (over 75s living alone). All of these seem intuitively appropriate. The low income indicators supplanted all other socioeconomic indicators.

Table 4 shows the model obtained when the income support and family credit variables were excluded from the candidate set. There was no endogeneity so the modelling was by weighted least squares regression. This model contains seven variables, none of which is related to poverty, although many of the census based socioeconomic indicators are surrogate measures of income and material disadvantage.

Table 5 shows the results of applying the two models to a notional sum of £500m, which is roughly the amount spent on acute services in Northern Ireland annually. The allocations produced using the crude and effective (age weighted) populations are also shown for comparative purposes. Because the size of a population has by far the greatest influence on its need for health care size, any formula of this kind will have an age­sex cost curve for acute hospital services in Northern Ireland to distribute funds for acute hospital services to the health and social services boards. The model components include access to general practice, hospital beds, general practices, residential and nursing homes, and geriatric beds and use of private beds. The supply variables that were significantly associated with use of inpatient services were access to hospital beds, general practices, residential and nursing homes, and geriatric beds and use of private beds. The two stage least squares equation for all specialties was significantly endogenous ($F_{ skeptic } = 2.28; P < 0.05$).

A parsimonious model (with five variables) retained most of the explanatory power of the full model (adjusted $R^2 = 52\%$, table 3). This risk adjustment model has been adopted for use in conjunction with an age-sex cost curve for acute hospital services in Northern Ireland to distribute funds for acute hospital services to the health and social services boards. The formula comprises two income related variables, two health variables, and a "social fabric" variable (over 75s living alone). All of these seem intuitively appropriate. The low income indicators supplanted all other socioeconomic indicators.

Table 4 shows the model obtained when the income support and family credit variables were excluded from the candidate set. There was no endogeneity so the modelling was by weighted least squares regression. This model contains seven variables, none of which is related to poverty, although many of the census based socioeconomic indicators are surrogate measures of income and material disadvantage.

Table 5 shows the results of applying the two models to a notional sum of £500m, which is roughly the amount spent on acute services in Northern Ireland annually. The allocations produced using the crude and effective (age weighted) populations are also shown for comparative purposes. Because the size of a population has by far the greatest influence on its need for health care size, any formula of this kind will have an age­sex cost curve for acute hospital services in Northern Ireland to distribute funds for acute hospital services to the health and social services boards. The model components include access to general practice, hospital beds, general practices, residential and nursing homes, and geriatric beds and use of private beds. The supply variables that were significantly associated with use of inpatient services were access to hospital beds, general practices, residential and nursing homes, and geriatric beds and use of private beds. The two stage least squares equation for all specialties was significantly endogenous ($F_{ skeptic } = 2.28; P < 0.05$).

A parsimonious model (with five variables) retained most of the explanatory power of the full model (adjusted $R^2 = 52\%$, table 3). This risk adjustment model has been adopted for use in conjunction with an age-sex cost curve for acute hospital services in Northern Ireland to distribute funds for acute hospital services to the health and social services boards. The formula comprises two income related variables, two health variables, and a "social fabric" variable (over 75s living alone). All of these seem intuitively appropriate. The low income indicators supplanted all other socioeconomic indicators.

Table 4 shows the model obtained when the income support and family credit variables were excluded from the candidate set. There was no endogeneity so the modelling was by weighted least squares regression. This model contains seven variables, none of which is related to poverty, although many of the census based socioeconomic indicators are surrogate measures of income and material disadvantage.

Table 5 shows the results of applying the two models to a notional sum of £500m, which is roughly the amount spent on acute services in Northern Ireland annually. The allocations produced using the crude and effective (age weighted) populations are also shown for comparative purposes. Because the size of a population has by far the greatest influence on its need for health care size, any formula of this kind will have an age­sex cost curve for acute hospital services in Northern Ireland to distribute funds for acute hospital services to the health and social services boards. The model components include access to general practice, hospital beds, general practices, residential and nursing homes, and geriatric beds and use of private beds. The supply variables that were significantly associated with use of inpatient services were access to hospital beds, general practices, residential and nursing homes, and geriatric beds and use of private beds. The two stage least squares equation for all specialties was significantly endogenous ($F_{ skeptic } = 2.28; P < 0.05$).

A parsimonious model (with five variables) retained most of the explanatory power of the full model (adjusted $R^2 = 52\%$, table 3). This risk adjustment model has been adopted for use in conjunction with an age-sex cost curve for acute hospital services in Northern Ireland to distribute funds for acute hospital services to the health and social services boards. The formula comprises two income related variables, two health variables, and a "social fabric" variable (over 75s living alone). All of these seem intuitively appropriate. The low income indicators supplanted all other socioeconomic indicators.

Table 4 shows the model obtained when the income support and family credit variables were excluded from the candidate set. There was no endogeneity so the modelling was by weighted least squares regression. This model contains seven variables, none of which is related to poverty, although many of the census based socioeconomic indicators are surrogate measures of income and material disadvantage.

Table 5 shows the results of applying the two models to a notional sum of £500m, which is roughly the amount spent on acute services in Northern Ireland annually. The allocations produced using the crude and effective (age weighted) populations are also shown for comparative purposes. Because the size of a population has by far the greatest influence on its need for health care size, any formula of this kind will have an age­sex cost curve for acute hospital services in Northern Ireland to distribute funds for acute hospital services to the health and social services boards. The model components include access to general practice, hospital beds, general practices, residential and nursing homes, and geriatric beds and use of private beds. The supply variables that were significantly associated with use of inpatient services were access to hospital beds, general practices, residential and nursing homes, and geriatric beds and use of private beds. The two stage least squares equation for all specialties was significantly endogenous ($F_{ skeptic } = 2.28; P < 0.05$).

A parsimonious model (with five variables) retained most of the explanatory power of the full model (adjusted $R^2 = 52\%$, table 3). This risk adjustment model has been adopted for use in conjunction with an age-sex cost curve for acute hospital services in Northern Ireland to distribute funds for acute hospital services to the health and social services boards. The formula comprises two income related variables, two health variables, and a "social fabric" variable (over 75s living alone). All of these seem intuitively appropriate. The low income indicators supplanted all other socioeconomic indicators.

Discussion

This study represents a considerable advance on previous work on risk adjustment because we used direct measures of poverty at small area level rather than indirect census based proxies. It is widely acknowledged that understanding of the association between socioeconomic standing, health status, and the need for health services would be enhanced if data directly reflecting income levels were more readily available.23 In addition, four of the five variables in our model (including household income) can be updated between censuses. This is clearly important for a formula used to allocate resources on an annual or three yearly basis. Our work is also an improvement on the current formula used in England in the following respects: more precise cost data were available; there was accurate and current measurement of access to private beds in health service hospitals; and the effect of distance from acute beds was empirically estimated by specialty.

The previous British government’s decision to damp down the effect of the “York formula” on allocations in the English NHS caused some controversy.24 This decision limited the extent of transfer of resources from the shire counties to metropolitan districts. It is notable, therefore, that the main effect of our formula that included social security benefits was to move resources from the board centred on Belfast to those serving primarily rural parts of Northern Ireland.
Our study shows the potential for using data on poverty to develop risk adjustment formulas so long as care is taken to identify the appropriate data and to separate out the relations between supply and demand. Peer review of formulas used by government is a new development but is also essential to assure local populations that scarce resources are being shared equitably.

We thank Stephanie Harcourt, Karen Campbell, David Marshall, Stephen Donnelly, and Sandy Fitzpatrick for providing the data and helpful advice.

Contributors: JQJ initiated the study and he and RAC-H were responsible for its overall design. DO'R and MRS were responsible for assembling the data. DO'R and JR developed the spatial interactive modelling methods and undertook the modelling. MRS and BM undertook the 2SLS modelling. All authors contributed to interpretation of the study results, with RAC-H and DO'R making the major contributions. JQJ and RAC-H wrote the first draft of the paper and revised it with DO'R and MRS. All the authors approved the final version. JQJ is the guarantor.

Funding: Northern Ireland Department of Health and Social Services.

Competing interests: JQJ has received research funds for a member of staff from the Northern Health and Social Services Board. RC-H is self-financing and carrying out the study meant that there were sufficient funds to pay his salary.