Use of regression analysis to explain the variation in prescribing rates and costs between family practitioner committees

D P FORSTER

C E B FROST

SUMMARY. There are proposals to set up prescribing budgets for family practitioner committees (now family health services authorities) and indicative prescribing amounts for practices. An intelligible model is therefore required for specifying budgetary allocations. Regression analyses were used to explain the variation in prescription rates and costs between the 98 family practitioner committees of England and Wales in 1987. Fifty one per cent of the variation in prescription rates and 44% of the variation in prescription costs per patient could be explained by variations in the age–sex structure of family practitioner committees. The standardized mortality ratio for all causes and patients in 1987, and the number of general practice principals per 1000 population in 1987, but not the Jarman under-privileged area score were found to improve the predictive power of the regression models significantly (P<0.01). The predictions of the model for the 10 family practitioner committees with the highest and lowest prescription rates or costs are reported and discussed. Potential improvements in models of prescribing behaviour may be thwarted by two problems. First, the paucity of readily available data on health care need at family practitioner committee and practice levels, and secondly, the increasing complexity in the statistical techniques required may render the procedure less intelligible, meaningful and negotiable in a contentious field.

Introduction

AMONG the many proposals for the reform of the National Health Service are prescribing budgets for family practitioner committees (now family health services authorities) and indicative prescribing budgets for general practices.[1] There has been debate as to whether these proposed budgets imply cash limits, thus raising the issue of whether doctors will have to consider even more carefully the prescription of effective yet costly drugs for their patients. Moreover, there has been a concern that the proposals may provide a disincentive for doctors to screen patients for diseases such as hypertension, which, if detected, could lead to drug treatment.[2] The response of the Department of Health has been to soften the language used and the word 'budget' is employed only in the context of the indicative prescribing scheme for family practitioner committees and the word 'amount' is used instead for practices.[3]

Putting aside the medico-political debate concerning the advisability or otherwise of such budgets, a key question is whether data and methods exist to specify prescribing budgets for family practitioner committees and the equivalent for practices. Unless one accepts the view that each general practitioner throughout the country should be issuing the same number of prescriptions per patient or have identical prescribing costs per patient, a model is required linking the factors most likely to account for variations in prescribing behaviour with observed variations in such behaviour.

The purpose of this study was to explore one method of predicting or explaining observed variations in prescription rates and costs between family practitioner committees in terms of 'need for health care', including age–sex structure, and indicators of resource availability.

D P Forster, FFCM, senior lecturer in community medicine, University of Newcastle Medical School. C E B Frost, BSc, lecturer in economics, School of Management and Business Studies, University of Sheffield.
Submitted: 19 January 1990; accepted: 29 August 1990.

Method

All data were collected for the 98 family practitioner committees of England and Wales. The total number of prescriptions in all therapeutic classes and the total costs of prescriptions for 1987 were derived from a representative one in 200 sample of prescriptions written by general practitioners and dispensed by retail pharmacies.[4] Similar prescription and cost data from dispensing general practitioners were collated and added to those dispensed by retail pharmacies.[4] The population data used were the Office of Population Census and Surveys mid-year estimates for each family practitioner committee in 1987.[5]

Three indicators of 'need for health care' were considered: the standardized mortality ratio for all causes and patients in 1987 (compares each family practitioner committee with a baseline of 100 for England and Wales as a whole); the Jarman under-privileged area score; and the age–sex structure of family practitioner committees in 1987. The Jarman score is a composite indicator based on eight social and demographic factors weighted by the importance placed on them for determining workload by a random sample of general practitioners.[6] The Jarman score is derived from data in the last decennial census in 1981. The indicator of resource availability considered was the supply rate of general practitioners in terms of the number of general practice principals per 1000 population in 1987.[7]

Ordinary least squares linear regression analysis was used to predict the number of prescriptions per patient for each of the 98 family practitioner committees, that is to say, regression analysis was used to establish the nature of the relationship between two or more variables.[8] In this study, there are two or more sets of values (for example, the prescription rate, the standardized mortality ratio and other demographic variables) and the value of the prescription rate that would correspond with given values of the standardized mortality ratio and other variables is predicted. The accuracy of the prediction will depend upon the strength of the relationship. This strength is usually measured in terms of the $\bar{R}^2$ statistic with a maximum value of 1.0 indicating 'a very strong relationship' and 0.0 indicating 'no discernable relationship'. In practice, less extreme values are frequently encountered. The same analytical method was used to estimate the value of total prescription costs per person (the predicted variable) given the values of the predictor variables (for example, the standardized mortality ratio and demographic variables).

Results

Figure 1(a) summarizes the results when simple linear regression was carried out using the standardized mortality ratio as the predictor variable and the prescription rate in 1987 as the variable to be predicted. In this analysis 46% of the variation in prescription rates between the family practitioner committees was explained by the variation in standardized mortality ratio ($\bar{R}^2$ = 0.46, P<0.001). A similar procedure was used with the Jarman score as the predictor variable and in this instance 16% of the variation in the prescription rate was explained by the variation in the Jarman score ($\bar{R}^2$ = 0.16, P<0.001) (Figure 1(b)).

These preliminary regressions suggest that it might be possible to model prescribing behaviour at family practitioner committee level, but they also indicate that there is considerable room for improvement. One plausible interpretation of the standardized mortality ratio regression is that there is some combination of age–sex effects in operation, and that such effects might be better investigated separately. Hence the regression results reported in Appendix 1 distinguish demographic factors from the remaining indicators of 'need for health care'. A selective summary of these results is reported in Table 1. From Table 1 it may be inferred that a superior model of prescribing behaviour will involve demographic factors in addition to the other two indicators of 'need for health care'. Overall, 51% of the variation in the prescription rate between the 98 family practitioner committees was explained by demographic factors alone. The addition of the standardized mortality ratio, the Jarman score and supply rate of general practitioners improved the predictive power of the model to 65%. No obvious advantage, however, was conferred upon the model by the inclusion of the Jarman score in the set of predictor variables (Appendix 1). In a second set of regression models, the same approach was used to explain the variation in prescribing costs per patient. In this case 44% of the variation in prescribing costs between the 98 family practitioner committees was explained by demographic variables alone. The inclusion of the standardized mortality ratio, Jarman score and supply rate of general practitioners offered a considerable improvement in that 60% of the overall variation was explicable. Once again the inclusion of the Jarman score in the group of predictor variables conferred no clear advantage (Appendix 1).

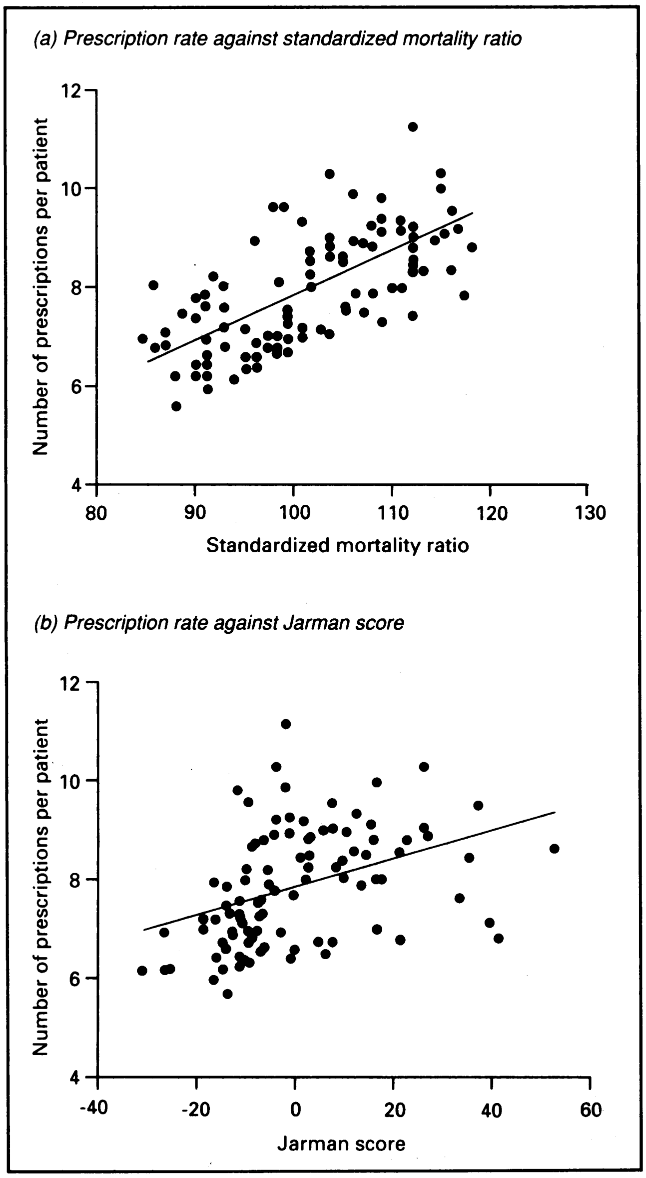

Figure 1. *Prescription rate for each family practitioner committee in 1987 against (a) standardized mortality ratio and (b) Jarman underprivileged area score.*

Table 1. Prediction of prescription rates and costs for all therapeutic classes.

Predictor variables	$\bar{R}^{2a}$ for: No. of prescriptions per patient	Cost of prescriptions per patient
Simple regressions		
Standardized mortality ratio	0.46	N/A
Jarman score	0.16	N/A
Phase I regressions		
Age–sex structure	0.51	0.44
Phase II regressions		
Phase I plus SMR, Jarman score, no. of GPs per 1000 population	0.65	0.60

[a]$\bar{R}^2$ represents the proportion of the variation in prescription rates between the 98 family practitioner committees which is accounted for by the predictor variable or variables. N/A = not applicable. SMR = standardized mortality ratio.

Table 2 sets out the 10 family practitioner committees with the highest and lowest actual prescribing rates in 1987. These ranged from 11.1 prescriptions per person in Mid Glamorgan to 5.7 in Oxfordshire. These observed prescribing rates are contrasted with the rates predicted by the regression model described in Appendix 1 (phase II model). Table 2 also sets out the extremes in prescribing costs per patient in 1987. These varied from £54 per person in Mid Glamorgan to £35 in Oxfordshire, although the relative positions of other family practitioner committees for actual prescription rates and costs were not identical. The overall figures for England and Wales are also given in Table 2. It is important to note that for the family practitioner committees with the highest actual prescribing rates and costs, the predicted value tends to underestimate the actual value's departure from the implied regression line, and vice versa for the family practitioner committees with the lowest rates and costs. The spread of the differences between the actual and predicted prescription rates was more evenly distributed for the family practitioner committees with less extreme rates and costs.

Table 2. Actual and predicted prescribing rates and costs in all therapeutic classes for the 10 family practitioner committees with the highest and lowest actual values in 1987.

	No. of prescriptions per patient			Cost of prescriptions per patient (£)	
Family practitioner committee	Actual	Predicted[a]	Family practitioner committee	Actual	Predicted[a]
Mid Glamorgan	11.1	9.2	Mid Glamorgan	54.1	48.0
West Glamorgan	10.4	8.9	North Tyneside	52.9	50.1
Liverpool	10.4	9.8	Salford	52.7	49.5
Salford	10.1	9.7	Gwynedd	51.9	47.8
Gwent	9.9	8.5	West Glamorgan	51.7	47.1
Barnsley	9.8	8.1	Gwent	49.8	45.2
Trafford	9.6	7.7	Liverpool	49.6	49.7
Gwynedd	9.6	9.0	Barnsley	49.4	43.9
Manchester	9.5	9.2	Clwyd	49.2	44.2
Sandwell	9.3	9.5	Trafford	48.9	42.4
Barnet	6.5	6.7	Hertfordshire	37.0	39.5
Gloucestershire	6.5	7.3	Wiltshire	36.7	38.7
Croydon	6.4	6.7	Enfield and Haringey	36.5	39.8
Wiltshire	6.4	6.9	Berkshire	36.4	36.6
Hertfordshire	6.3	7.1	Gloucestershire	36.2	41.7
Bromley	6.3	6.9	Bromley	36.0	39.3
Surrey	6.2	6.8	Croydon	35.9	37.4
Berkshire	6.2	6.6	Greenwich and Bexley	35.9	39.5
Buckinghamshire	6.0	6.1	Buckinghamshire	35.9	35.1
Oxfordshire	5.7	5.9	Oxfordshire	35.0	35.2
England and Wales	7.7	—	England and Wales	41.9	—

[a]By regression model described in Appendix 1.

Discussion

The approach to explaining the variation in prescription rates used in this study has been to accept that general practitioners' responses to morbidity in their practices will, in part, be met by prescriptions. Thus, explanations of prescribing behaviour are only understandable and likely to be successful if the level of morbidity in a practice or family practitioner committee area can be measured. An indicator of 'need for health care', or more accurately 'need for a prescription' is elusive. There is a paucity of useful sources of meaningful morbidity data at general practice or family practitioner committee level. For example, continuous morbidity recording in general practice is carried out by relatively few general practitioners.[9]

For this reason, three proxy indicators of the need for health care were considered: the standardized mortality ratio, the Jarman score and demographic factors, that is, age–sex structure. The all causes standardized mortality ratio, used first, may appear inappropriate since much general practice work does not involve life threatening conditions.[10] It does, however, have the advantage of being readily available each year at family practitioner committee level and could be obtained at individual practice level on a regular basis, probably by aggregating data for several years and using, say, a five year rolling average. It is noteworthy that, even after allowing for the effect of the demographic factors, the standardized mortality ratio was found to exert a positive influence on the prescription rate and costs.

The second indicator of health care need used was the Jarman underprivileged area score derived from 1981 census data. This indicator is intended to convey the pressure of work on general practitioners. Given that the eight elements making up the Jarman score are weighted by the subjective views of general practitioners with respect to the pressure of work generated, it was anticipated that the Jarman score might give a good explanation of the variation in prescription rates. Regression equations using the Jarman score as a predictor variable, however, gave a relatively poor explanation of the variation in prescription rates or costs. Furthermore, the inclusion of the Jarman score along with other indicators of 'need for health care' did not improve the explanatory power of the models. It seems likely that the Jarman score, based essentially on 1981 data, may be inapplicable for use in years distant from the date of the census. Moreover, the Jarman score would be impossible to calculate accurately from census data for individual practices.[11,12]

In Appendix 1, variation in resource availability in the form of the number of general practitioners per 1000 population was included in the regression models for the following reasons. In most studies of the use of health services, it has been found that increased availability of supply increases health service utilization.[13,14] Elsewhere, it has been argued that the availability of good medical care varies inversely with the need for health care.[15] Apparently, the supply rate of general practitioners does offer a worthwhile improvement when employed as a predictor variable. In addition, the supply rate appears to be positively related to the prescription rate and costs. This finding appears to corroborate the previously noted utilization effects and is not inconsistent with the 'inverse care' law.[15]

For all practical purposes a regression model explaining about 65% of the observed variation in prescription rates and 60% in the case of costs is available if the standardized mortality ratio and supply rate of general practitioners are combined with demographic measures. Ideally, it would be preferable to specify a regression model explaining a much higher percentage of the observed variation in prescription rates and costs. There are a number of ways in which such an improvement might be attained. First, additional predictor variables could be introduced into the linear regression equations. Secondly, it may be possible to improve the explanation of prescribing behaviour by using different regression analysis techniques. Briefly, these might include weighted least squares regression or non-linear least squares regression methods.

The government intends that family practitioner committees with above average actual prescribing costs should bring their expenditure into line with lower spending family practitioner committees. It is accepted, however, that family practitioner committees may have high prescribing costs because of special known factors. In our analysis, we have interpreted these special factors principally as 'need for health care' including demographic factors. We have introduced a reasonably satisfactory model of prescribing behaviour, which is noteworthy for its reliance on demographic factors, the standardized mortality ratio and the supply rate of general practitioners. One practical implication of this model is that family practitioner committees with high prescribing costs per patient, such as Mid Glamorgan or North Tyneside, would come under less financial pressure than if they were compared with some nationally derived average. In practice, since family practitioner committee budgets will probably be decided at regional level, each family practitioner committee's actual prescription rate and costs will be compared with the appropriate regional average. Whether there would be concomitant generosity in the allocation of prescribing budgets to family practitioner committees in the south, for example, Buckinghamshire and Oxfordshire, which are underspending relative to the model's predictions is a moot point.

In family practitioner committees in which prescription rates or prescribing costs are in excess of those predicted by the model, two explanations are possible. It may be that prescribing is unjustifiably high, or that the high prescription rate or costs reflect factors not contained in the model. We favour the latter explanation and have noted that routinely available data do not exist to improve the predictive power of the model. We have indicated that there are a number of technical and statistical improvements that might be made to the model presented here, but at some risk of increasing its overall complexity. In which case the model of prescribing behaviour becomes less understandable, meaningful and negotiable. This bodes ill in a field in which the ideology of budgets has not generally been well received.

While we agree with the government's concentration on age and sex structure in setting family practitioner committee budgets and indicative prescribing amounts in practices, our view is that the scheme is premature, given that some 49% of prescribing variation is unexplained.

Appendix 1

The regressions reported here were undertaken in two phases. From studies in selected practices, it is known that morbidity rates and prescription rates vary in different age and sex groups, and are particularly high in the young and elderly age groups.[9] Using available demographic data about the proportions of a family practitioner committee's population in the young or elderly age bands and the reproductive age group for women, maximum prediction of the observed variation in prescription rates or costs was sought from the phase I regressions. More precisely, in statistical terms, the square of the product-moment correlation between the dependent variate and the set of independent variates, known as the coefficient of determination, adjusted for degrees of freedom, that is $\bar{R}^2$, was maximized.[16]

At the conclusion of phase I, the overall variation in the prescription rate or costs was divided into two parts: the part predicted or explained by the age and sex distributions in the family practitioner committees and the part remaining to be explained (the residual variation). In phase II, therefore, other potentially relevant indicators, for example the Jarman score and standardized mortality ratio, were introduced into the regression analysis to see if their effects on prescription rates or costs were statistically significant. In addition, the resource availability indicator, the supply rate of general practitioners, was introduced at this stage.

Table 3 summarizes the phase I results for prescription rates for all therapeutic classes, and for prescribing costs. In the usual case, the regression coefficient of a predictor variable measures the effect on the predicted variable of a unit increase in the predictor variable, holding any other predictor variables constant. When regressions involve proportions, the interpretation given to the regression coefficients must be amended because the proportion of children under one year of age cannot be increased, while at the same time holding the remaining age–sex proportions constant. In these regressions, the coefficients of the regression may be related to the prescription rate for a particular age–sex group.[16] For example, for the prescription rate for children aged under one year:

Regression coefficent of constant + regression coefficient of children aged under one year = 6.5 + 383.9 = 309.4

This interpretation immediately raised a question as to why certain coefficients were statistically significant and yet had unrealistically large or even negative values. There probably exists a strong linear relation connecting many, if not all, the predictor variables. The predictor variables

Table 3. Phase I regressions: explanation of the variation in prescription rates and prescribing costs for all therapeutic classes (dependent variables) by the age–sex structure of the 98 family practitioner committees in 1987.

Label	Partial regression coefficient	t	Significance level
No. of prescriptions per patient [a]			
Constant	6.45	1.45	NS
Proportion of population:			
Aged under 1 year	383.85	2.37	P<0.05
Aged 1–4 years	−67.91	−1.32	NS
Males aged 75 years and over	−291.97	−7.15	P<0.001
Females aged 65–74 years	207.92	6.03	P<0.001
Females aged 15–44 years	−18.18	−1.30	NS
Prescribing costs per patient [b]			
Constant	20.31	6.80	P<0.001
Proportion of population:			
Males aged 75 years and over	−1047.49	−6.54	P<0.001
Females aged 65–74 years	923.42	8.79	P<0.001

In each of the regressions reported above, $\bar{R}^2$ was maximized using various elements of family practitioner committee age–sex structure. NS = not significant. [a]92 degrees of freedom. $\bar{R}^2$ = 0.51. [b]95 degrees of freedom. ($\bar{R}^2$) = 0.44.

Table 4. Phase II regressions: explanation of the variation in prescription rates and prescribing costs for all therapeutic classes (dependent variables) by the indicators of 'need for health care' of family practitioner committees in 1987.

Label	Partial regression coefficient	t	Significance level
No. of prescriptions per patient [a]			
Constant	2.63	0.52	NS
Proportion of population:			
Aged under 1 year	67.52	0.43	NS
Aged 1–4 years	−9.07	−0.20	NS
Males aged 75 years and over	−158.98	−2.95	P<0.01
Females aged 65–74 years	115.94	3.37	P<0.01
Females aged 15–44 years	−26.74	−2.05	P<0.05
No. of GPs per 1000 population	6.28	2.95	P<0.01
Jarman score	0.01	1.46	NS
SMR (all causes and patients)	0.05	3.17	P<0.01
Prescribing costs per patient [b]			
Constant	−21.06	−2.46	P<0.05
Proportion of population:			
Males aged 75 years and over	−432.84	−1.90	NS
Females aged 65–74 years	586.64	4.51	P<0.001
No. of GPs per 1000 population	33.85	3.86	P<0.001
Jarman score	−0.003	−0.11	NS
SMR (all causes and patients)	0.26	4.08	P<0.001

In each of the regressions reported above, $\bar{R}^2$ was initially maximized using family practitioner committee age–sex structure (see Table 3). The need for health care indicators including the supply rate of general practitioners were subsequently added. NS = not significant. SMR = standardized mortality ratio. [a]89 degrees of freedom. $\bar{R}^2$=0.65. [b]92 degrees of freedom. $\bar{R}^2$ = 0.65.

are then said to be collinear and the coefficients of the regression become indeterminate with occasionally, but not inevitably, large standard errors.[16]

Age–sex standardized prescription rates and costs were required for phase II regressions involving the standardized mortality ratio and possibly the Jarman score. The phase I regressions provided a means of obtaining such data indirectly, and thus the collinearity problem did not matter given the strong relationship between the predicted variable and the set of phase I predictors.

The effects of the remaining 'need for care' and resource availability indicators on prescription rates and the attendant prescribing costs were explored by a number of regressions. In each case the conclusion was virtually the same, the additional explanatory power offered by the inclusion of these variables, when added separately or together, was substantial except in the case of the Jarman score. A representative set of these phase II results is given in Table 4. Although the standardized mortality ratio and the Jarman score were apparently collinear, their inclusion in the same regression equation did not materially affect the phase II results reported here.

References

1. Secretaries of State for Health, Wales, Northern Ireland and Scotland. *Working for patients: indicative prescribing budgets for general medical practitioners. NHS review working paper 4*. London: HMSO, 1989.
2. Spencer JA, van Zwanenberg TD. Prescribing research: PACT to the future. *J R Coll Gen Pract* 1989; **39:** 270-272.
3. Department of Health. *Improving prescribing. The implementation of the GP indicative prescribing scheme.* London: Department of Health, 1990.
4. Prescription Pricing Authority. *Statistical data relating to prescriptions dispensed by chemists, contractors and dispensing doctors for 1987. Forms PD1 and FP51.* London: Department of Health, 1987.
5. Office of Population Censuses and Surveys. *Key population and vital statistics, England and Wales for 1987. Series VS No.14/PPI No. 10.* London: HMSO, 1989.
6. Jarman B. Identification of underprivileged areas. *Br Med J* 1983; **286:** 1705-1709.
7. General Medical Services. *Basic statistics for 1987.* London: Department of Health, 1987.
8. Godfrey K. Statistics in practice: simple linear regression in medical research. *N Engl J Med* 1985; **313:** 1629-1636.
9. Royal College of General Practitioners, Office of Population Censuses and Surveys and Department of Health and Social Security. *Morbidity statistics from general practice. Third national study, 1981–82.* London: HMSO, 1986.
10. Fry J. *Common diseases.* Lancaster: MTP Press, 1985.
11. Hutchinson A, Foy C, Smyth J. Providing census data for general practice. 1. Feasibility. *J R Coll Gen Pract* 1987; **37:** 448-450.
12. Foy C, Hutchinson A, Smyth J. Providing census data for general practice. 2. Usefulness. *J R Coll Gen Pract* 1987; **37:** 451-454.
13. Feldstein MS. Effects of differences in hospital bed scarcity on type of use. *Br Med J* 1964; **2:** 562-565.
14. Cullis JG, Forster DP, Frost CEB. The demand for inpatient treatment: some recent evidence. *Applied Economics* 1980; **12:** 43-60.
15. Hart JT. The inverse care law. *Lancet* 1971; **1:** 405-412.
16. Maddala GS. *Econometrics.* New York: McGraw Hill, 1977.

Acknowledgements

We are grateful to the Department of Health for access to data and to Karen Cowley for secretarial assistance.

Address for correspondence

Dr D P Forster, University of Newcastle upon Tyne, School of Health Care Sciences, Division of Community Medicine, The Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH.