How much variation in clinical activity is there between general practitioners? A multi-level analysis of decision-making in primary care

Peter Davis, Barry Gribben[1], Roy Lay-Yee[2], Alastair Scott[3]

Department of Public Health and General Practice, Christchurch School of Medicine and Health Sciences, University of Otago, Christchurch; [1]Department of General Practice and Primary Health Care, [2]Department of Community Health, [3]Department of Statistics, University of Auckland, Auckland, New Zealand

Objectives: **There is considerable policy interest in medical practice variation (MPV). Although the extent of MPV has been quantified for secondary care, this has not been investigated adequately in general practice. Technical obstacles to such analyses have been presented by the reliance on ecological small area variation (SAV) data, the binary nature of many clinical outcomes in primary care and by diagnostic variability. The study seeks to quantify the extent of variation in clinical activity between general practitioners by addressing these problems.**

Methods: **A survey of nearly 10 000 encounters drawn from a representative sample of general practitioners in the Waikato region of New Zealand was carried out in the period 1991–1992. Participating doctors recorded all details of clinical activity for a sample of encounters. Measures used in this analysis are the issuing of a prescription, the ordering of a laboratory test or radiology examination, and the recommendation of a future follow-up office visit at a specified date. An innovative statistical technique is adopted to assess the allocation of variance for binary outcomes within a multi-level analysis of decision-making.**

Results: **As expected, there was considerable variability between doctors in levels of prescribing, ordering of investigations and requests for follow up. These differences persisted after controlling for case-mix and patient and practitioner attributes. However, analysis of the components of variance suggested that less than 10% of remaining variability occurred at the practitioner level for any of the measures of clinical activity. Further analysis of a single diagnostic group – upper respiratory tract infection – marginally increased the practitioner contribution.**

Conclusions: **The amount of variability in clinical activity that can definitively be linked to the practitioner in primary care is similar to that recorded in studies of the secondary sector. With primary care doctors increasingly being grouped into larger professional organisations, we can expect application of multi-level techniques to the analysis of clinical activity in primary care at different levels of organisational complexity.**

Journal of Health Services Research & Policy Vol 7 No 4, 2002: 202–208

Introduction

Geographic variations in rates of hospitalisation and of medical and surgical intervention are well documented.[1] The bulk of research to date has focused on the secondary and tertiary sectors of care, much of it relating to particular specialties[2] or to hospital treatments more generally.[3] Less work has been carried out on variability in patterns of clinical activity in primary care,[4] although analysis of variations in activity has been growing for purposes such as the derivation of capitation-based resource allocation formulae.[5]

Substantial inter-practitioner variation in patterns of primary care activity has been established for over a decade for a wide range of systems and across a diversity of measures. Such differences persist for radiology, tests, prescriptions, referrals, follow-up and night visits.[6] Variations in patient charges,[7] return visit intervals[8] and hospital admission rates[9] have also been the subject of empirical research.

The theoretical and methodological dimensions of work on medical practice variation (MPV) have become increasingly sophisticated. Explicit hypotheses have been developed that draw on differences in population morbidity rates, levels of supply and the appropriateness of clinical decisions; key explanatory concepts have been those of practice organisation, professional norms, supplier-induced demand, professional uncertainty and physician practice style. In theoretical terms, this amounts to a 'supply' hypothesis; that is, an explanatory

Peter Davis PhD, Professor of Public Health, Department of Public Health and General Practice, Christchurch School of Medicine and Health Sciences, University of Otago, PO Box 4345, Christchurch, New Zealand. **Barry Gribben MBChB**, Senior Research Fellow, Department of General Practice and Primary Health Care, **Roy Lay-Yee MA**, Data Analyst, Department of Community Health, **Alastair Scott PhD**, Professor of Statistics, Department of Statistics, University of Auckland, New Zealand.

Correspondence to: PD.

framework giving causal primacy to the role of practice and practitioner attributes in the functioning of the health care system, particularly in accounting for MPV.[6]

One of the key deficiencies in the literature has been the reliance on data drawn from small area variations (SAV).[1,10] Although SAV data provided the key initial insights into MPV, they are also clearly susceptible to the ecological fallacy. This has greatly hampered the drawing of any clear policy implications from MPV research.[11]

As a correction to the unit of analysis problem,[12] some investigators have concentrated on micro-level studies into the decision-making processes of individual practitioners on a limited number of clinical problems using simulated, imaginary or real-life cases.[13] Although generating important findings at the level of clinical decision-making, these studies have limited generalisability and, with some exceptions, do not necessarily incorporate contextual effects.[14]

The growing acceptance of multi-level statistical techniques, together with the availability of patient-level data, has now made it possible validly to estimate parameters and attribute variation at different levels of aggregation. These techniques have been applied in an increasing range of studies in primary care, including prescribing[15] and resource use.[16]

A further issue in the analysis of data of this kind on MPV has been the binary nature of much of the information available. Frequently, the crucial outcomes are dichotomous. This is particularly so in the analysis of medical decision-making and patterns of clinical activity. Essentially, the investigator is presented with information on whether or not a practitioner performed some intervention, such as writing a script or ordering an investigation. This has limited the kind of analyses that can be carried out. In particular, it has not been possible to allocate the proportion of variance attributable to the practitioner level using a hierarchical model of analysis. However, important advances have been made recently in developing an extension of R^2 to the binary case.[17] The procedure assumes that a dichotomous outcome is determined by an underlying threshold model (see Methods section for further details).

Finally, it should be noted that there is the potential for inter-practitioner variation in clinical activity to be masked by the degree of diagnostic variability in general practice. It has been argued that variations in recorded diagnostic rates are mainly due to 'the consistent but idiosyncratic and selective exclusion by practitioners of some components from the total set which often coexists in a new diagnosis'.[18] Among the conditions apparently least susceptible to this effect were respiratory problems.

The aim of this paper is to quantify the relative influence of practitioner characteristics on variations between family doctors in patterns of clinical activity. It seeks to do so by applying an innovative technique to the analysis of binary data and by focusing a subanalysis on a specific respiratory condition in the quest to control diagnostic variability.

Methods

Study details

The study was carried out over the period September 1991 to August 1992 in the Waikato region of New Zealand. Centred on the provincial city of Hamilton, the region has an ethnically mixed, urban–rural population of 320 000. In demographic terms, the Waikato region can be said to provide a representative cross-section, but not a replica, of the country as a whole.[19] The general practitioner (GP) community in the Waikato is also reasonably representative,[20] although data for the years 1989–1990 for GP availability (population per full-time equivalent GP) and for levels of utilisation (consultations per capita) show the Waikato to be slightly above the national average on both counts.[21]

As in the rest of New Zealand, GP incomes in the Waikato are almost entirely determined by direct charges to patients on a fee-for-service basis – unless covered by private insurance – supplemented to a limited extent by public subsidy and by social insurance cover for medical treatment following personal injury by accident.[22]

The data for this study are drawn from a survey of general practice encounters. Encounters were selected in a two-stage process designed to generate a 1% sample of all general practice consultations in the Waikato region. Eighty per cent of all practitioners took part in the first phase of data collection. A decline in compliance followed the first phase of data collection and, overall, data collection was successfully completed in 69% of all possible participating doctor–weeks (collection phases). Data are presented here for those 143 doctors completing at least ten encounter forms; this amounted to 9746 records, representing nearly 85% of all encounters collected in the survey.

Variables

For each consultation, data were collected on patient demographics – these included age group, ethnicity and gender – and on the diagnosis of any problems identified by the practitioner at the encounter. These constitute 'case-mix' variables that are potential confounders to be controlled, and are presented as binary items, as follows:

- Diagnosis: practitioner-identified problems coded into 16 International Classification of Primary Care (ICPC) chapters.[23]
- Patient demographics: age, ethnicity (pakeha (European), other) and gender (male/female).

In a previous survey, practitioners were also asked to provide information on themselves and their practices, including age, workload and practice size.[19] These provide a broader set of binary 'supply' variables in the form of the following practitioner and practice characteristics:

- Whether or not the doctor was in a group practice (defined as one with three or more doctors).
- The age group of the doctor (whether or not the doctor was over 45 years old).

- Practitioner's workload (whether or not the doctor provided more than seven of a possible ten consulting sessions per week).

The three clinical activity variables were constructed as binary outcomes:

- Prescribing: whether or not a script was issued at the conclusion of the encounter.
- Test ordering: whether or not an investigation was ordered.
- Follow-up: whether or not an arrangement was made for follow-up at a specified later date.

Building on earlier work theoretically informed by the 'supply' hypothesis,[6] three analytically important determinants – income incentives, physician agency and clinical ambiguity – were operationalised and measured as follows:

- Income incentives (local doctor density): each practice was allocated to one of six strata according to the patient:practitioner ratio for the area.
- Physician agency (encounter initiation): the practitioner was asked to indicate on the encounter form who had initiated the visit (doctor, patient, other).
- Clinical ambiguity (diagnostic uncertainty): practitioners were asked to indicate on the encounter form the level of uncertainty associated with the main diagnosis for the visit (by ticking one of the categories 'none', 'low' or 'high').

These three variables are treated as factors to be controlled because of their potential significance as background influences on inter-practitioner variability.

Descriptive statistics for all variables are reported in Table 1. It should be noted that the data are presented for encounters. Only the top five diagnostic groups are reported since these accounted for nearly 80% of the total.

Statistical analysis

Because of the hierarchical nature of the data, with patients being sampled from practitioners' weekly workloads, cluster effects are likely.[12] Furthermore, the patient subsamples varied in size from 12 to 120 between doctors. It was therefore convenient to use pooled data for instances in which these practitioner subsamples were small and potentially unstable. Multi-level or hierarchical modelling addresses the analytical issues raised by these sorts of data – random effects, hierarchical data structures and efficient parameter estimates across subsamples of varying sizes – and was used in the analysis.[24] Using the notation of MLwiN,[25] the analytical model is described as follows:

Let p_{ij} denote the probability that the outcome of interest – say, that a script will be written – will occur at the ith encounter by the jth doctor. Our logistic model for p_{ij} takes the form:

$$\text{logit}(p_{ij}) = \log\frac{p_{ij}}{1-p_{ij}} = \beta_{0j} + \beta_1 A_{ij} + \beta_2 A_{ij}^2 + \beta_3 G_{ij} + \beta_4 E_{ij} + \beta_5 \text{icpca}_{ij} + \ldots + \beta_{20}\text{icpcy}_{ij} + \beta_{21}\text{numgps}_j + \beta_{22}\text{docage}_j + \beta_{23}\text{fulltime}_j + \beta_{24}\text{dens}_j + \beta_{25}\text{init}_{ij} + \beta_{26}\text{uncert}_{ij} + \beta_{27}(\text{dens}_j \times \text{uncert}_{ij})$$

where: b_{0j} is an intercept specific to the jth doctor (this is a normally distributed random variable, i.e. $b_{0j} = b_0 + \mu_{0j}$, with a mean β_0 and a variance $\sigma^2_{\mu 0}$); A_{ij}, G_{ij} and E_{ij} are variables for age, gender (1 = 'male') and ethnicity (1 = 'European'); $\text{icpca}_{ij} \ldots \text{icpcy}_{ij}$ are dummy variables representing each of the 16 chapters of the ICPC diagnostic system classifying the problem (or problems) of the patient at the ith encounter with the jth doctor; 'numgps', 'docage' and 'fulltime' are variables for the number of doctors in the practice (1 = three or more doctors), the doctor's age group (1 = over 45 years) and doctor workload (1 = more than 7/10) for the jth doctor; the last four terms in the model are dummy variables representing the analytical variables 'dens' (doctor density), 'init' (doctor initiated), 'uncert' (clinical uncertainty), with 'dens × uncert' a product term to measure the 'doctor density' by 'clinical uncertainty' interaction, all at the ith encounter with the jth doctor.

By specifying that only the intercept in the model is random, we have assumed that the effect of each of the variables in the model is the same for each doctor. The model can be visualised as a series of parallel regression lines, one for each doctor.[26]

Table 1 Dataset

Variables	Percentage[a] (n = 9746)
Diagnostic group[b]	
Respiratory	21.8
Musculoskeletal	17.5
Skin	15.1
General	12.7
Cardiovascular	11.3
Patients' attributes	
Average age (years)	35.3
Male	42.2
European	80.8
Practitioner variables	
Doctor aged over 40 years	44.3
Doctor in full-time practice[c]	50.8
Fewer than three doctors in practice	40.6
Clinical activity rates	
Prescribing	62.0
Test ordering	12.4
Follow-up	78.1
Analytical variables	
High doctor density	61.5
Doctor-initiated visit	12.9
High diagnostic uncertainty	7.4

[a]Except for patient age, percentage of encounters in which the stated attribute was present.
[b]The diagnosis groups are International Classification of Primary Care chapter headings for the top five groups reported.
[c]'Full-time' means working at least eight-tenths.

Snijders and Bosker[17] have proposed an extension of the definition of R^2 suggested by McKelvey and Zavoina for a single-level logistic model that allows the proportion of variance explained by a multi-level random intercept logistic model to be calculated. The procedure assumes that, as above, the dichotomous outcome Y_{ij} is determined by an underlying threshold model and conceives the variance of the underlying variable to be composed of the variance of the linear predictor, the variance of the intercept and the level one residual variance for a logistic model, the variance of the logistic distribution ($\pi^2/3$). The variance of the linear predictor, σ_F^2, is calculated as the variance of the estimated values for Y_{ij} when only the fixed terms of the model are used in the prediction calculation. Thus the total variance of the underlying variable, Y_{ij}, can be written as:

$$\mathrm{Var}(Y_{ij}) = \sigma_F^2 + \sigma_{\mu 0}^2 + \left(\frac{\pi^2}{3}\right)$$

The proportion of variance explained by the model is then:

$$\frac{\sigma_F^2}{\sigma_F^2 + \sigma_{\mu 0}^2 + (\pi^2/3)}$$

and the proportion of variance at level 2 (the intraclass correlation) is:

$$\frac{\sigma_{\mu 0}^2}{\sigma_{\mu 0}^2 + (\pi^2/3)}$$

The MLwiN software package was used.[25] Models were fitted using second-order predictive quasi-likelihoods where convergence could be achieved; otherwise – as in the case of the analysis of the subset of respiratory disorders – a simpler first-order marginal quasi-likelihood was estimated.[24]

Results

The results of applying the basic model to the data are displayed in Table 2. It can be seen that there are small, but consistent, changes in the variance components across all three activity measures with the addition of diagnosis and then the remainder of the variables.

It is interesting to note that the variance at the practitioner level increased for all three measures on the introduction of diagnosis into the fixed part of the model. This can happen with certain distributions of non-random confounders. There are various plausible interpretations, but no information is available to further these possible explanations.

An outline of the variance components generated in this analysis is presented in Table 3. The proportion of variance explained by the full model varied from just under a third for prescribing, to a fifth for investigations, and over 15% for follow-up. Of the residual variance, the proportion at the doctor level varied from less than 5% to just over 10%.

In order to address the issue of diagnostic variability and to control case-mix more exactly, all encounters for which there was a single diagnosis of upper respiratory tract infection were identified and the analysis repeated. The results are reported in Table 4. The consequence of rendering the analysis more homogeneous by diagnosis is that the proportion of residual variance at the doctor level increases, in one instance markedly.

The results in Table 4 should be viewed with caution since the first-order marginal quasi-likelihood approximation used to calculate them can produce biased estimates under some conditions.[28] Although improved methods are available,[29] we could not get these methods to converge here. It should be noted that for all models there was no evidence of extra-binomial variation. This means that the use of $\pi^2/3$ as an estimate of patient variance was justified.

Discussion

Findings

Research into MPV in primary care has been hampered by several technical and conceptual difficulties. In the first place, the use of SAV data has risked the ecological fallacy by drawing undue inferences from aggregated data. One solution to this has been to resort to highly focused studies (for example, using clinical vignettes) at the practitioner level. However, this in turn fails to capitalise on important contextual information. Multi-level statistical techniques permit the full use of data at a number of levels of aggregation,[30] and this has been deployed successfully in the current study.

Secondly, an important dimension to work of this kind is the analysis of decision-making in the clinical setting. Typically decisions of this kind in medical practice are binary in nature. This has presented considerable technical difficulties in the allocation of variance between levels of aggregation,[17] but this has been overcome to a reasonably satisfactory degree in the current investigation.

Table 2 Multilevel analysis[a] of clinical activity ($n = 9746$)

	Prescribing		Investigations		Follow-up	
	Intercept (β_0)	Variance ($\sigma^2_{\mu 0}$)	Intercept (β_0)	Variance ($\sigma^2_{\mu 0}$)	Intercept (β_0)	Variance ($\sigma^2_{\mu 0}$)
Null model	0.466 (0.028)[b]	0.144 (0.023)	−1.983 (0.041)	0.295 (0.049)	1.294 (0.042)	0.474 (0.052)
+ diagnosis	−0.766 (0.065)	0.181 (0.028)	−2.261 (0.086)	0.315 (0.054)	0.730 (0.074)	0.483 (0.053)
+ all variables	−0.579 (0.139)	0.141 (0.027)	−2.741 (0.197)	0.260 (0.051)	0.825 (0.181)	0.406 (0.062)

[a]MLwiN estimation method: iterative generalised least squares using second-order predictive quasi-likelihoods.
[b]Standard error in parentheses.

Table 3 Analysis[a] of variance components: full model ($n = 9746$)

	Prescribing	Investigations	Follow-up
% encounters	62.0	12.4	78.1
Variance of linear predictor (fixed part)	1.41	0.94	0.66
Residual variance (random part)	3.431	3.55	3.696
At doctor level	0.141 (0.027)[b]	0.26 (0.051)	0.406 (0.062)
At patient level ($\pi^2/3$)	3.29	3.29	3.29
Total variance	4.841	4.49	4.356
% total variance explained by linear predictor	29.1	20.9	15.2
% residual variance at doctor level	4.1	7.3	11.0
% residual variance at patient level	95.9	92.7	89.0

[a]MLwiN estimation method: iterative generalised least squares using second-order predictive quasi-likelihoods.
[b]Standard error in parentheses.

Table 4 Analysis[a] of variance components: upper respiratory tract infection, single diagnosis ($n = 465$)

	Prescribing	Investigations	Follow-up
% encounters	77.4	9.0	65.4
Variance of linear predictor (fixed part)	1.139	0.900	0.193
Residual variance (random part)	3.776	4.522	3.796
At doctor level	0.486 (0.259)[b]	1.232 (0.578)	0.506 (0.218)
At patient level ($\pi^2/3$)	3.29	3.29	3.29
Total variance	4.915	5.422	3.989
% total variance explained by linear predictor	23.2	16.6	4.8
% residual variance at doctor level	12.9	27.2	13.3
% residual variance at patient level	87.1	72.8	86.7

[a]MLwiN estimation method: iterative generalised least squares using first-order marginal quasi-likelihoods.
[b]Standard error in parentheses.

Finally, there is the issue of inter-practitioner variability in diagnostic judgements. Just as practitioners may follow certain 'styles' of clinical decision-making, so they may do so in their patterns of diagnosis.[18] Again, this is a problem that others have attempted to address using standardised clinical vignettes.[13] This solution is relatively artificial and distanced from the real-life practice setting, however. Instead, we identified a problem cluster (upper respiratory tract infection) that is likely to be subject to a lower than average influence from both patient and practitioner variability in presenting and diagnostic behaviour, respectively.

Shortcomings

There are a number of methodological shortcomings to the current investigation that need to be taken into account in evaluating the results. In the first place, there is the regional character of the study – together with its less than resounding response rate – which potentially limits the generalisability of its findings. Although the population and practitioner communities are reasonably representative of the New Zealand norm,[20] the region may not have included a sufficiently large and diverse demographic and health service base to provide an adequate range of variation.

Secondly, the operationalisation of the key variables in the model may be imperfect. Among the many potential measures of clinical activity in general practice that could have been used in an analysis of this kind, only three are reported here. However, these three items – test orders, prescribing and request for follow-up – represent a substantial commitment of resources in general practice (apart, that is, from the practitioner's own time). A fourth possible candidate for inclusion – referral – could not be used in this analysis because of the very small numbers of encounters in which referrals were made.

Finally, the model fitted – around variation in the intercept – assumed that variation between practitioners was restricted to variations in the underlying rate of the measured clinical activity. A more sophisticated model – permitting variation in slopes – would allow for the possibility that practitioners varied in their response to the presence of one or more confounding variables. These models are computationally considerably more complex and difficult to fit than the model fitted in this paper. A number of more complicated models were unsuccessfully explored, due to problems with convergence and the stability of the obtained estimates.

Interpretation

Other studies of inter-practitioner variation in clinical activity in the hospital setting have established a range of between 2%[31] and 10%[32] attributable to physician attributes. In keeping with these earlier investigations, this study found that something in the region of 10% of all variability in decision-making outcomes – after controlling for relevant case-mix and background factors – could be attributed to the doctor level. Is this 'significant' in policy and practice terms? If, following a Bayesian logic, our prior expectation was that there should be no clinically significant variability between doctors after controlling for relevant factors, then the fact that there appears to be a weighting of '10%' may appear important. If, on the other hand, our prior

expectation – say, from the MPV and SAV literature – was that the impact of systems factors on variability in clinical activity was considerable and of major policy significance, then a finding of '10%' variability due to doctors may appear relatively unimportant.

Marmot[33] has argued that this 'individual variance' approach understates the importance of key causative variables of policy interest. Thus, in his study of mortality in a cohort of British civil servants, grade of employment together with age only explained 2.2% of variation in coronary heart disease, despite the fact that the lowest grade had a 70% increased risk compared to the highest. As for mortality from lung cancer in the same cohort, smoking, age, respiratory function and other risk factors explained only 7.4% of variability. Yet smoking is the major cause of lung cancer.

The individual-level variability in this study represents the outcomes of a host of measured and unmeasured factors across a sample of patients represented in their full diversity. Few of these are evidently susceptible to policy or public health intervention. By contrast, the doctor-level variability reflects the outcomes of the work of a far more homogeneous group – that is, professionals operating in well-defined practice settings and trained within the established parameters of orthodox medical practice. The '10%' recorded in this study, therefore, may not give a straightforward indication as to the weight in policy and practice that should be accorded to the role of doctor variability.

Conclusions

The literature on MPV has established some propositions. Firstly, a clear distinction has to be made between the conclusions that can be drawn from ecological versus multi-level data. Secondly, variability is likely to be particularly marked in areas of greater professional uncertainty.[34] Thirdly, there are influences on MPV that can be detected for recognisable practitioner attributes such as gender, years in practice and diagnostic style.[35] Fourthly, as confirmed by this study, there remains a significant level of inter-practitioner variability in clinical activity, even after controlling for standard case-mix and organisational and professional factors. Finally, a more or less stable clinical 'style' can be confirmed across a range of activities.[35,36] This study has indicated that the amount of doctor-specific variability may be less substantial than previously inferred from ecological studies and that the opportunity for guiding change through practitioner interventions may be less than previously thought. Nevertheless, the techniques outlined here bring primary care into the established field of assessing institutional performance which, until now, has been largely applied to the hospital sector, particularly in surgery.[37] Despite the apparently limited practitioner effects demonstrated in this study, it is also the case that primary care doctors are increasingly being grouped into larger professional organisations. Therefore, we can expect a more sustained and systematic extension of these techniques to the analysis of clinical activity in the primary sector at a range of organisational levels.

Acknowledgements

We are very grateful to the nearly 200 GPs of the Waikato region who gave so willingly of their time to make the collection of the data for this study possible. We wish also to thank the members of the WaiMedCa Trust Committee chaired by Dr Linda Rademaker for their understanding and cooperation in facilitating access to the data.

References

1. Folland S, Stano M. Small area variations: a critical review of propositions, methods, and evidence. Medical Care Review 1990; 47: 419–465
2. Birkmeyer JD, Sharp SM, Finlayson SR, Fisher ES, Wennberg JE. Variation profiles of common surgical procedures. Surgery 1998; 124: 917–923
3. Ashton CM, Petersen NJ, Souchek J, Menke TJ, Yu HJ, Pietz K et al. Geographic variations in utilization rates in Veteran Affairs hospitals and clinics. New England Journal of Medicine 1999; 340: 32–39
4. Parchman ML. Small area variation analysis: a tool for primary care research. Family Medicine 1995; 27: 272–276
5. Rice N, Dixon P, Lloyd DCEF, Roberts D. Derivation of a needs based capitation formula for allocating prescribing budgets to health authorities and primary care groups in England: regression analysis. BMJ 2000; 320: 284–288
6. Davis P, Gribben B, Scott A, Lay-Yee R. The 'supply hypothesis' and patterns of clinical activity in general practice. Social Science and Medicine 2000; 50: 407–418
7. Woodward CA, Hutchison B, Norman GR, Brown JA, Abelson J. What factors influence primary care physicians' charges for their services? An exploratory study using standardized patients. Canadian Medical Association Journal 1998; 158: 205–207
8. Schwartz LM, Woloshin S, Wasson JH, Renfrew RA, Welch HG. Setting the revisit interval in primary care. Journal of General Internal Medicine 1999; 14: 230–235
9. Reid FD, Cook DG, Majeed A. Explaining variation in hospital admission rates between general practices: cross sectional study. BMJ 1999; 319: 98–103
10. Paul-Shaheen P, Clark JD, Williams D. Small area analysis: a review and analysis of the North American literature. Journal of Health Politics, Policy and Law 1987; 12: 741–809
11. Stano M. Further issues in small area variations analysis. Journal of Health Politics, Policy and Law 1991; 16: 573–588
12. Divine GW, Brown JT, Frazier LM. The unit of analysis error in studies about physicians' patient care behavior. Journal of General Internal Medicine 1992; 7: 623–629
13. McKinlay JB, Burns RB, Feldman HA, Freund KM, Irish JT, Kasten LE et al. Physician variability and uncertainty in the management of breast cancer. Results from a factorial experiment. Medical Care 1998; 36: 385–396
14. Clark JA, Potter DA, McKinlay JB. Bringing social structure back into clinical decision making. Social Science and Medicine 1991; 8: 853–866
15. Davis P, Gribben B. Rational prescribing and interpractitioner variation. A multilevel approach. International Journal of Technology Assessment in Health Care 1995; 11: 428–442
16. Scott A, Shiell A. Analysing the effect of competition on general practitioners' behaviour using a multilevel modelling framework. Health Economics 1997; 6: 577–588
17. Snijders TAB, Bosker RJ. Multilevel analysis. An introduction to basic and advanced multilevel modelling. London: Sage, 1999

18. Crombie DL, Cross KW, Fleming DM. The problem of diagnostic variability in general practice. Journal of Epidemiology and Community Health 1992; 46: 447–454
19. McAvoy B, Davis P, Raymont A, Gribben B. The Waikato Medical Care (WaiMedCa) survey 1991–1992. New Zealand Medical Journal 1994; 107 (suppl part 2): 387–433
20. Gribben B, Bonita R, Broad J, McAvoy B, Raymont A. Geographical variations in the organisation of general practice. New Zealand Medical Journal 1995; 108: 361–363
21. Malcolm L. Trends in primary medical care related expenditure in New Zealand 1983–1993. New Zealand Medical Journal 1993; 106: 470–474
22. Brown MC, Crampton P. New Zealand policy strategies concerning the funding of general practitioner care. Health Policy 1997; 41: 87–104
23. Lamberts H, Woods M, eds. ICPC. International classification of primary care. Oxford: Oxford University Press, 1987
24. Goldstein M. Multilevel statistical models. London: Edward Arnold, 1995
25. Goldstein H, Rabash J, Plewis I, Draper D, Browne W, Yang M et al. A user's guide to MLwiN. London: Institute of Education, University of London, 1998
26. Duncan C, Jones K, Moon G. Context, composition and heterogeneity: using multilevel models in health research. Social Science and Medicine 1998; 46: 97–117
27. de Leeuw J, Kreft IGG. Software for multilevel analysis. In: Leyland AH, Goldstein H, eds. Multilevel modelling of health statistics. Chichester: Wiley, 2001: 187–204
28. Rodriguez G, Goldman N. An assessment of estimation procedures for multilevel models with binary responses. Journal of the Royal Statistical Society, Series A 1995; 158: 73–89
29. Goldstein H, Rasbash J. Improved approximations for multilevel models with binary responses. Journal of the Royal Statistical Society, Series A 1996; 159: 505–513
30. Blakely TA, Woodward AJ. Ecological effects in multi-level studies. Journal of Epidemiology and Community Health 2000; 54: 1–7
31. Hayward RA, Manning WG, McMahon LF, Bernard AM. Do attending or resident physician practice styles account for variations in hospital resource use? Medical Care 1994; 32: 788–794
32. Roos NP. Hospitalization style of physicians in Manitoba: the disturbing lack of logic in medical practice. Health Services Research 1992; 27: 361–384
33. Marmot M. Social causes of social inequalities in health. Working paper series no. 99.01. Cambridge, MA: Harvard Center for Population and Development Studies, 1999
34. Bronstein JM, Cliver SP, Goldenberg RL. Practice variation in the use of interventions in high-risk obstetrics. Health Services Research 1998; 32: 825–839
35. Franks P, Williams GC, Zwanziger J, Mooney C, Sorbero M. Why do physicians vary so widely in their referral rates? Journal of General Internal Medicine 2000; 15: 163–168
36. Davis P, Gribben B, Scott A, Lay-Yee R. Do physician practice styles persist over time? Continuities in patterns of clinical decision-making among general practitioners. Journal of Health Services Research & Policy 2000; 5: 200–207
37. Clare Marshall E, Spiegelhalter DJ. Institutional performance. In: Leyland AH, Goldstein H, eds. Multilevel modelling of health statistics. Chichester: Wiley, 2001: 127–142