Medical Education 1986, **20**, 181–186

Admission for medicine in the United Kingdom: a structural model of background factors

I. C. McMANUS† & P. RICHARDS‡

†*Department of Psychiatry, St Mary's Hospital Medical School and Department of Psychology,*
University College London, and ‡*Department of Medicine, St Mary's Hospital*
Medical School, London

Summary. A statistical model is presented of the direct and indirect influences of 21 variables upon success at selection for medical school.

Key words: *School admission criteria; *Models, theoretical; Educational status; Social class; Sex factors; London

Introduction

The process of selection for admission to medical school is complex, and several aspects of it have been described by us in a series of papers (McManus & Richards 1984a, b, c, 1985; McManus 1985), in which we examined all applications to St Mary's Hospital Medical School, London, during the autumn of 1980 for entry in October 1981. In one of the studies (McManus & Richards 1984a) a hierarchical logistic regression analysis was used to determine the influence of 24 background variables: *demographic* features, such as age, sex, social class and family background; *educational* factors such as O- and A-level results, size and type of school, etc.; and *applicational* factors, such as date of application, choices on Universities Central Council on Admissions (UCCA) form, etc. It was found that the only significant predictors of success, defined as entry to *any* of the medical schools to which the student had applied, were the number and the grade of A-levels, the number and the grade of O-levels, the date of UCCA application, and whether or not the applicant was from a medical family (defined as either parent being medically qualified). In addition, a subsequent analysis (McManus & Richards 1985) also showed that applicants with a non-European surname (NES) were significantly less likely to be accepted. In our analyses we commented (p.1204) that although factors such as social class and education had no direct influence upon success at application, they did have indirect effects by affecting factors such as A-level success which are themselves important in selection. This paper reports a comprehensive structural analysis of background factors, using a covariance modelling technique, and demonstrates both direct and indirect influences on admission, and on the factors preceding and associated with it.

Method

The interrelationships between 22 variables were examined in a total of 986 applicants to St Mary's Hospital Medical School, for whom adequate information was available. The Pearsonian correlation matrix between all variables was modelled by means of the LISREL IV program (Joreskog 1978). In view of the different measurement scales of variables, standardized coefficients were analysed throughout, and hence it was the correlation rather than a covariance matrix which was modelled. More detailed descriptions of the distributions of the variables may be found in McManus & Richards (1984a). Binary variables, such as sex,

Correspondence: Dr I. C. McManus, Department of Psychiatry, St Mary's Hospital, Praed Street, London W2 1NY, England.

were entered into the correlation matrix in the same fashion as other variables (i.e. as *phi* correlations). In consequence there are minor differences between the present analysis and that reported in McManus & Richards (1984a), where logistic regression was used: none of the differences are substantial. A few variables in the analysis of McManus & Richards (1984a) were omitted from the present analysis since it was not clear how to enter them into the causal model.

The structural model

The variables in the analysis were ordered into an *a priori* arrangement which was felt to be compatible with known causal constraints between the variables. Thus O-level grade was interpreted as causally prior to A-level grade since the correlation between O-level and A-level grades might reasonably be regarded as high O-level grades causing (in some sense) high A-level grades, but not vice versa. For the variables shown in Fig. 1 the hierarchical rule used was that a variable could be caused by any variable to the left of it, and could cause any variable to the right of it. The five variables on the left-hand margin were regarded as beyond further causal analysis and represented, in the LISREL terminology, the exogenous or X variables, the observed interrelations between them being included in the *phi* matrix. All other variables were endogenous or Y variables, and causal interrelations between them were included in the asymmetric beta matrix, and influences upon them from the X variables were included in the gamma matrix. A few of the Y variables are shown vertically one above the other, and for these there was no reasonable *a priori* causal ordering, and adirectional correlations between them were included in the *psi* matrix. Two relations should have been regarded as causal according to the above schema (school size—private education; private education—sixth-form size) but were actually represented by adirectional correlations in the *psi* matrix, since the direction of causation was not sufficiently clear.

In general, it was not difficult to decide on an *a priori* ordering of the variables, since events preceding other events in time could usually be regarded as potential causal influences on subsequent variables. Thus, for instance, the decision as to how many O-levels are taken must be prior to the grades actually obtained in those O-levels. In other cases causality could be

Figure 1. Shows a structural model of variables having direct and indirect influences upon applications to and acceptance by medical schools. Direct causal effects are shown by single-headed straight arrows moving in a left-to-right direction. Correlations without any causal direction are shown by curved double-headed arrows. Standardized structural coefficients are indicated alongside arrows. All coefficients are significant with $P<0{\cdot}10$; those with a single underlining are significant with $P<0{\cdot}05$, and those with a double-underlining are significant with $P<0{\cdot}001$. Structural effects with negative coefficients are indicated by dashed lines. It should be noted that since higher social class and earlier date of UCCA application are indicated by *lower* numbers, then the sign of structural influences upon and by them should be interpreted with care. The five X-variables shown down the left-hand side influence many variables, and for clarity no attempt has been made to draw direct causal arrows from them; rather abbreviations have been placed close to the Y-variables which are influenced.

Detailed descriptions of variables. 'Social class' (abbreviated to 'C' within body of figure): Registrar-General's social class; 'Medical family' (abbreviated to 'Med'): one or more parents medically qualified; 'NES': applicant has a non-European surname; 'Sex': applicant is female; 'North': applicant lives in the north of England, or in Scotland or Northern Ireland (see McManus & Richards, 1984a, for definition); 'Private Ed'n': applicant educated for one or more years at a non-State school (i.e. a 'public' school); 'School size', 'Sixth form', and 'University entry': total number of students in the school, the number in the sixth form, and the number going to university each year; 'N O-levels', 'N A-levels': total number of O(A) levels taken; 'O-level grade', 'A-level grade': average grade attained at O(A) level, awarding 5 points for an A, 4 for a B, etc.; 'Maths', 'Biology': Maths (biology) taken at A-level; 'Oxb'ge': Oxford or Cambridge included on the UCCA form; 'UCCA date': date at which application received at UCCA; 'Brack.': amount of bracketing used on UCCA form, high scores indicating *no* brackets, and lower scores indicating all choices shown as first equal (see McManus & Richards, 1984a, for details); 'N Lond', 'N E&W', 'N S&NI': number of medical schools in London, England and Wales (excluding London and Oxford and Cambridge), and in Scotland and Northern Ireland included on the UCCA form; 'Accept': applicant accepted at *any* medical school to which they had applied.

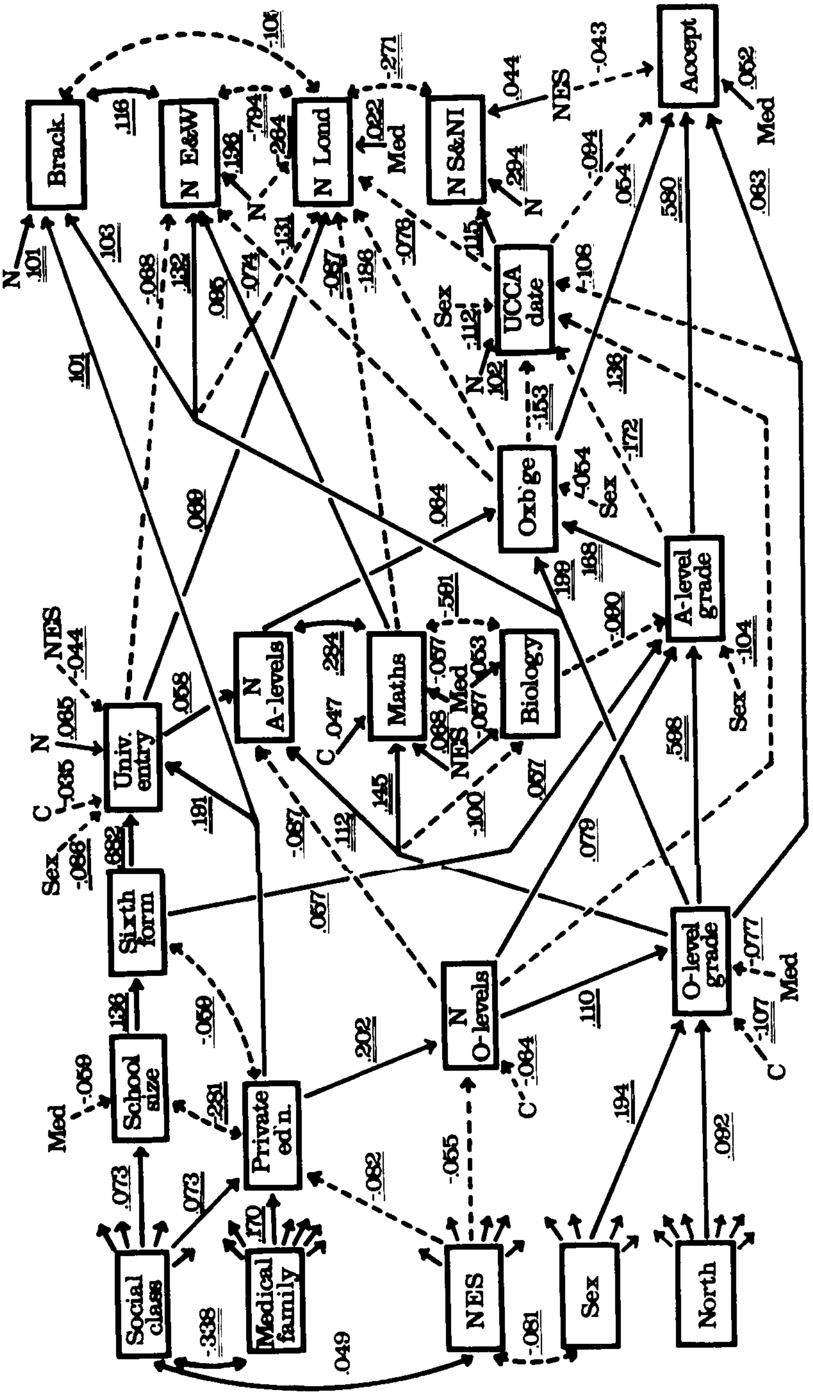
Social class
Medical family
NES
Sex
North
School size
Private ed'n.
Sixth form
Univ. entry
N O-levels
O-level grade
N A-levels
Maths
Biology
A-level grade
Oxb'ge
UCCA date
Brack.
N E&W
N Lond
N S&NI
Accept
Med
Sex
N
C
NES

Fig. 1

inferred from the progress from general to specific; thus the number in a school going to university might reasonably determine the number of A-levels that a particular student might take, while the reverse causal process is inconceivable, both in the sense that the influence of the group upon the individual will be larger than that of the individual upon the group, and that the persons going to university are in a previous year to those who are deciding to take A-levels, and hence have already been accepted by the time that the next group is deciding how many A-levels to take.

The variables on the UCCA form are not quite so obvious in their ordering. For a post-A-level candidate, the decision to apply to Oxbridge will be influenced by A-level grades, and will itself cause an earlier application to UCCA. All other aspects of the UCCA form will probably be subsequent to these two, although that is by no means certain. A pre-A-level candidate will have a certain ability at the time of application to UCCA which can be regarded as a latent variable which causes both the Oxbridge application and the subsequent A-level grade; the consequence will be almost identical to the post A-level case for modelling purposes, although in the presence of more information about pre-A-level performance the two cases could be distinguished.

It is inevitable that there will be grey areas in the creation of a complex structural model, and these might appear as being arbitrary assumptions. Interested readers may satisfy themselves as to their importance or otherwise by fitting alternative models to the correlation matrix (which is available on request from the authors).

A note on structural modelling

The aim of structural modelling is to describe causal and non-causal interrelations between a series of models. The data for analysis consist of the correlation or covariance matrices between all possible pairs of variables. An adequate model is defined as one for which the differences between the actual elements of the matrix and those predicted by the model (the residuals of the elements) are no greater than would be predicted as a result of chance variation, and a *chi*-square statistic is available to test the adequacy of the fit. Relations between variables in the model can be of two forms: causal, indicated in path analytic diagrams by straight arrows with a single head, indicating that A causes B, and not vice versa; and non-causal, indicated by curved, double-headed arrows, indicating an association between A and B for which no causal mechanism is postulated. Substantially different path models produce different predictions for the correlation matrix, and hence their relative adequacies as descriptions of the data may be tested.

The most convenient method of fitting a causal model involves the computer program LISREL, for which eight matrices must be specified to define a particular model. Since the current model has no latent, or unmeasured, variables, the two lambda matrices in the LISREL specification can be set as identities. Variables are defined as exogenous (only influencing other variables) or endogenous (influenced by other variables). In the present analysis, social class, medical family, NES, sex and being from northern Britain are exogenous variables. Causal relations between endogenous variables appear in the beta matrix, causal relations from exogenous to endogenous variables appear in the gamma matrix, non-causal relations between endogenous variables appear in the *psi* matrix, non-causal relations between exogenous variables appear in the *phi* matrix. Error variances for the endogenous variables appear in the diagonal of the *theta-epsilon* matrix, and of the exogenous variables in the *theta-delta* matrix. Any relation between variables can be fixed at zero, or can be allowed to be free, in which case its best value will be found by the program, using a maximum likelihood procedure which, in addition, allows estimation of the parameter's standard error.

More detailed accounts of causal modelling may be found in Kenny (1979), Cohen & Cohen (1983) and Long (1983).

Results

Initially, a fully saturated model was fitted in which the *gamma* matrix was completely free, the *beta* matrix was completely free between hierarchical levels, and the *psi* matrix was completely free within hierarchical levels. Parameter estimates which gave z-values of less than one were then forced to zero, and the model re-estimated. All parameters which were not then significant with $|z|>1{\cdot}645$ (i.e. $P<0{\cdot}10$) were then forced to zero, and a further model estimated in which all parameters were significant with $P<0{\cdot}10$. This several stage process was used to take account of the potentially confusing effects of multicollinearity on the significance of structural coefficients. All the parameter estimates shown in Fig. 1 are significant with $P<0{\cdot}10$, and the majority are much more significant than that. The model as shown is an adequate description of the entire correlation matrix (*chi*-squared for goodness of fit$=166{\cdot}64$, 1144 df, $P=0{\cdot}095$).

Discussion

This model as presented demonstrates the contention expressed in our previous paper that many factors can have an indirect influence on acceptance for medical school. Thus, although social class has no direct influence upon admission, those of higher social class tend to have higher O-level grades but take fewer O-levels, are less likely to take A-level maths, and go to private sector schools, to smaller schools, and to schools which have a higher university entry each year. Each of these factors subsequently affects other factors. Nevertheless, it may be seen from the estimates of effect size that in general the indirect influences are relatively small (they may be estimated by multiplying coefficients of paths connected in series and adding coefficients of paths connected in parallel—see Cohen & Cohen [1983]).

The complexity of the structural diagram in Fig. 1 emphasizes the well-known truism that social phenomena are inherently complex. Nevertheless, it also shows that they are not infinitely complex; in particular there are many potential causal relations between variables for which there is no adequate empirical necessity. Thus, to take an example at random, students who have taken A-level biology might be thought to be more likely to be accepted at medical school, whereas not only is there no evidence for such a direct link, there is actually evidence of indirect effects whereby those taking biology tend to obtain lower A-level grades and are thus less likely to be accepted. Figure 1 allows the unravelling of a large number of such relationships. As a more complex example, consider the question of whether women find it more or less easy than men to be accepted for medical school. Sex has no direct influence upon acceptance. However, women tend to obtain higher O-level grades, which increase the likelihood of acceptance (indirect influence$=0{\cdot}194\times0{\cdot}063=0{\cdot}012$), and to obtain lower A-level grades which lower the likelihood of acceptance (indirect influence$=-0{\cdot}104\times0{\cdot}580=-0{\cdot}060$), they tend to apply earlier to UCCA, thereby increasing their chances of acceptance (indirect influence$=-0{\cdot}112\times-0{\cdot}094=0{\cdot}011$), and tend not to apply to Oxbridge, decreasing the likelihood of acceptance (indirect influence$=-0{\cdot}054\times-0{\cdot}054=-0{\cdot}003$). In addition there are further indirect influences; women applicants tend not to have non-European surnames (indirect influence$=-0{\cdot}081\times-0{\cdot}043=0.003$), thereby increasing women's likelihood of acceptance. It is clear, therefore, that there is no simple answer to the deceptively simple question of whether sex affects the chances of acceptance; different indirect influences have opposing (and generally small) effects upon acceptance.

In interpreting Fig. 1 it must be remembered that the estimates of structural coefficients may only be valid for the particular group in which the data were collected—applicants to medical school. The estimates may well not be valid for other populations, such as the entire school population, although they might provide reasonable starting estimates for such parameters.

References

Cohen J. & Cohen P. (1983) *Applied Multiple Regression/Correlation Analysis for the Behavioural Sciences* (2nd edition). Pp. 356-60. Lawrence Erlbaum, Hillsdale, New Jersey.

Joreskog K.G. (1978) *LISREL IV. User's Guide*. National Educational Resources, Chicago.

Kenny D.A. (1979) *Correlation and Causality*. John Wiley, Chichester.

Long J.S. (1983) *Covariance Structure Models: an Introduction to LISREL*, Sage, London.

McManus I.C. (1985) *Medical students: origins, selection, attitudes and culture.* MD Thesis, University of London.

McManus I.C. & Richards P. (1984a) An audit of admission to medical school: 1. Acceptances and rejects. *British Medical Journal* **289,** 1201–4.

McManus I.C. & Richards P. (1984b) An audit of admission to medical school: 2. Short-listing and interviews. *British Medical Journal* **289,** 1288–90.

McManus I.C. & Richards P. (1984c) An audit of admission to medical school: 3. Applicants' perceptions and proposals for change. *British Medical Journal* **289,** 1365–7.

McManus I.C. & Richards P. (1985) Admission to medical school. *British Medical Journal* **290,** 319–20.

Received 18 *February* 1985; *editorial comments to authors* 15 *March* 1985; *accepted for publication* 7 *October* 1985