PRELIMINARY & INCOMPLETE

Determinants of Medical School Dropout Probabilities for the United Kingdom

Wiji Arulampalam[+,*], *Robin A. Naylor*[+], and *Jeremy P. Smith*[+]

25th May 2001

Keywords: Medical students, Student Dropout (Non-completion) Probabilities, Discrete time hazard, Limited duration model.

JEL Classification No.: J24, I2, C41

[+] Address for correspondence: Department of Economics, University of Warwick, Coventry, CV4 7AL, UK.
Email:, *wiji.arulampalam@warwick.ac.uk*, *robin.naylor@warwick.ac.uk*, *jeremy.smith@warwick.ac.uk*
Web-site: *http://www.warwick.ac.uk/fac/soc/Economics*

[*] Also, IZA, Bonn, Germany.

Acknowledgement: A number of people have given us invaluable help in generating the dataset: in particular, we thank John McClure and Peggy Paull at UCAS, John McNeill in MIS at Warwick, and staff at DfEE, HEFCE, HESA and the Schools Register. We acknowledge both the USR, as the original depositors, and the UK Data Archive for the use of the data-set SN:3456 Universities' Statistical Record. None of these individuals or organisations bears any responsibility for any of the analysis or interpretations presented in this paper.

Abstract.

From individual-level longitudinal data for two entire cohorts of medical students in UK universities, we analyse the probability that an individual student will ‘drop out’ of medical school prior to the successful completion of their studies. We examine the cohort of students enrolling for a medical degree at the start of the academic years 1985 or 1986. We find evidence that medical student completion is influenced by prior qualifications, sex, and age as well as by the characteristics of the medical school itself.

1. Introduction

Labour economics focuses chiefly on the analysis of decentralised labour markets in which equilibrium outcomes emerge from the interplay of the forces of supply and demand, where these forces are governed significantly by the price of labour. The labour market for medical practitioners in the UK, however, in important respects is better described by a planning model than by a model of a purely decentralised free market. Primarily, this is because of the nature of the necessary regulation in the market. In the UK, doctors are trained within medical schools funded and regulated by the government. In large part, this arrangement follows from the still predominantly public nature of medical provision under the auspices of the National Health Service. Currently in the UK, as in many other countries, there are major concerns regarding a growing shortage of medical doctors. The UK solution to this problem is perceived as requiring more efficient planning of the medical workforce.[1] The Third Report of the Medical Workforce Standing Advisory Committee (MWSAC) to the Secretary of State for Health bears testimony to this reliance on a planning approach and also makes plain how such an approach – to be effective and efficient – demands a wealth of detailed data and appropriate accompanying analysis. In part, the nature of this information requirement is a consequence of the lengthy training period and the consequent time lags between forecasting future supply and demand and the subsequent ‘production’ of the medical workforce. Uncertainties in demographic trends – which are themselves endogenously determined with the nature and level of medical provision – and in the evolution of medical technologies themselves compound the difficulties associated with forecasting and planning.

[1] This is a very different approach from that recently recommended from an analysis of the shortage of postgraduate students of Economics in the UK (see Machin and Oswald, 2000). In that case, the problem was attributed largely to a distortion in the relative price of labour.

The Third Report of the MWSAC proposed a number of measures to prevent what it described as 'the current significant imbalance' between demand and the domestic supply of doctors from becoming 'increasingly severe'. A main conclusion of the Report was that there should be a substantial increase (of about 1,000 per annum) in medical student intake, together with policies to ensure 'minimised levels of wastage from such courses, thereby increasing the proportion of entrants that qualify as doctors.' Of course, a significant incentive to reduce wastage rates lies in the high cost associated with medical training. One element of the proposed package of measures to address the issue of minimising wastage concerned changes in selection procedures of candidates for medical schools in order to 'obtain graduates with a wider range of skills and interests.' The related issue of selection and admission of students into medical schools in the UK has itself been the recent focus of significant political controversy and debate.

The development of strategies to minimise the medical student dropout rate should be conducted in the light of analysis of the determinants of completion and withdrawal behaviour by medical students. This has been rendered difficult in the UK by a lack of reliable data and analysis. The MWSAC acknowledge that there is significant uncertainty regarding estimates of medical school qualification rates. Estimates are typically based on the annual intake of students at each medical school and the number qualifying 5 years later. This is imprecise partly because of variations over time and across institutions in the number of students taking intercalated degrees, for example, and hence taking longer to qualify. Thus, it is not surprising that there is so much variation in the estimated dropout rates, which fluctuated (with no systematic trend) between 8% and 14% during the period 1986/7 and 1991/2 (see MWSAC, 3rd Report, p. 65). So severe is the information problem regarding the drop-out rate that the MWSAC 3rd Report acknowledges that all its analysis and recommendations rest on an estimated student wastage rate which is itself based on ‘some anecdotal evidence that drop out is falling’ (p. 25).

In the absence of firm data on medical school dropout rates, different studies have produced widely varying estimates of the average national rate. Parkhouse (1996) uses University Statistical Record (USR) and Higher Education Funding Council (HEFCE) data on medical school intake and numbers qualifying five years later and estimates an average UK drop-out rate of between 11.7% and 14.1%. McManus (1996) disputes these figures and cites survey evidence that the rate is around 7% or 8%, with about half of the students who drop out doing so for non-academic reasons (see McManus *et al*., 1995).

Previous data and accompanying analysis have been based typically either on aggregated (medical school-level) official data from the USR or HEFCE or on follow-up surveys of particular sub-samples of medical students. This mirrors the situation regarding the analysis of all UK university students across all subject areas. Johnes and Taylor (1990), for example, is a classic example of the state of the art of modelling student withdrawal rates[2] from university-level data. Very recently, however, (anonymised) individual student-level administrative data for full population cohorts of students have become available to researchers. These data contain rich information not only on the academic characteristics of students in UK universities (their courses, institutional affiliation, performance, reasons for leaving, accommodation, *inter alia*) but also on their personal, social, and prior educational characteristics. These data offer the prospect of much more precise estimates of dropout rates and of detailed analytical investigation of the factors associated with dropout behaviour. This should be of interest and relevance to the issue of examining the impact of changing selection procedures on medical students’ dropout probabilities (see, for example, Angel and Johnson, 2000). Smith and Naylor

[2] Amongst other phenomena such as degree performance and labour market outcomes.

(2001) have exploited the individual-level data to examine the determinants of the student drop-out rate across all UK university students on three or four year degrees (which excludes medical students), for students who commenced their studies at university in the Autumn of 1989. Given the public policy importance of the issue, in the current paper we focus exclusively on a detailed analysis of the determinants of the withdrawal probability of UK medical students.

The issue of the determinants of medical student dropout probabilities is important and topical for a number of reasons in addition to informing policy discussions on medical retention rates in the context of an increasing shortage of supply of medical doctors. For example, there has been a lively and high-profile political debate in the UK concerning the extent of accessibility of medical schools to students regardless of their social or school background. This has led to explicit recommendations to broaden access to undergraduate medical education (see Angel and Johnson, 2000). Predicting the likely impact of such policies on retention and progression is clearly an important issue. Furthermore, there is a related and ongoing debate concerning the extent to which previous educational qualifications affect medical student performance and progression (see, for example, McManus *et al.*, 1999). Finally, the Government has recently introduced a series of Performance Indicators for Higher Education Institutions in the UK, including an indicator that ranks institutions on the basis of completion rates. Interpreting league tables based on completion rates should be done against the backdrop of an understanding of which factors influence dropout rates. Furthermore, it is important to assess the statistical significance of rankings represented in such league tables. We offer some analysis of these issues in the current paper.

The rest of this paper is organised as follows. Our econometric model is presented in Section 2 along with a discussion of relevant issues regarding our model estimation procedure. Section 3 describes the important features of the dataset and the variables used in the analyses. This is followed by Section 4 where we present the main results of our analysis. Final section closes the paper with a discussion and further remarks.

2. Econometric model and issues

Analysis of student dropout behaviour has received much attention in the US, where one of the most influential theoretical explanations of student attrition is the path analyses model of Tinto (1975, 1987). This model suggests that the student's social and academic integration into the educational institution is the major determinant of completion, and identifies a number of key influences on integration. These include the student's family background, personal characteristics, previous schooling, prior academic performance, and interactions between students and with faculty. In an analysis of the dropout behaviour of all UK university students matriculating in 1989, Smith and Naylor (2001) find evidence to support both the hypothesis that student completion is influenced by the extent of prior academic preparedness and the hypothesis that social integration at university is important. Smith and Naylor (2001) use a simple probit model based on the incidence of a student dropping out of university.

In the current paper, our objective is to model the conditional probability that an individual will drop out of a medical degree program during some small time interval, conditional on not having dropped out up to that point: that is, the hazard rate. The underlying variable is the time spent on the program rather than calendar time. Unlike in conventional duration models, two specific characteristics of the program need to be accounted for in this case. First, the program duration is limited to five years. Because of this limited duration, the underlying continuous time duration variable will have a distribution that is continuous over the interval (0,5) and a discrete probability mass at the end point of 5 years. Second, the program cannot be completed before the end of five years. That is, the probability of successfully completing the program during the first five years is zero.[3]

We use the Cox's Proportional Hazards Model as a starting assumption. Given the above characterisation of the programme of studies, the hazard for individual i, $\theta_i(t)$, is parameterised as

$$\theta_i(t)=\lambda(t)\exp(\mathbf{x_i}'\boldsymbol{\beta}) \qquad (t < 5 \text{ years}) \qquad (1)$$

where $\lambda(t)$ is the baseline hazard at time t, $\mathbf{x}_i$ is the vector of characteristics for individual i (excluding the intercept constant), and β is the corresponding vector of unknown coefficients. Because of possible measurement errors in the recording of the date of dropout, for the purpose of the analysis presented here, the duration information has been re-coded in terms of whole years completed (this is discussed in more detail below). A recorded duration of t whole years therefore indicates duration on the continuous time-scale, between t-1 and t years. Hence, the probability of exiting by time t conditional on $\mathbf{x}_i$, given that the student was still on the program at time t-1 is given by

$$h_{1i}(t \mid \mathbf{x}_i) = Prob[T_i < t \mid t-1 \le T_i] = 1 - exp\left\{-\int_{t-1}^{t} \theta_i(\tau) d\tau\right\}$$

$$= 1- exp\left[-\int_{t-1}^{t} \lambda(\tau)\, exp\{\, \mathbf{x}_i'\boldsymbol{\beta}\} d\,\tau\right]$$

$$= 1- exp\ \left[-exp\{\mathbf{x}_i'\boldsymbol{\beta} + \delta(t)\}\right] \qquad (2)$$

[3] Mealli, et. al. (1996) consider a duration-limited competing risks model in the context of Youth Training Programmes. Booth and Satchell (1995) and van Ours and Ridder (2000) look at a related model for PhD completion rates.

where $\delta(t) = ln\left\{\int_{t-1}^{t} \lambda(\tau)\, d\tau\right\}$ (3)

We thus have an Extreme-value form for the hazard model in discrete time.[4]

As seen earlier, the degree program ends at the end of year 5.[5] There is thus a forced termination at this point. In order to account for this, it is assumed that these limit point probabilities take the same Extreme-value form as before but with a different set of coefficients. This is specified as

$$h_{2i}\left(t=5\,|\,\mathbf{x}_i\right) = Prob[\text{dropping out in year 5} \mid \text{survival up to year 5, } \mathbf{x_i}]$$

$$= 1 - exp\left[-exp\left\{\mathbf{x}_i{}'\boldsymbol{\alpha} + \eta\right\}\right] \quad (4)$$

where $\mathbf{x}_i$ is the vector of characteristics for individual i (excluding the intercept constant), α is the corresponding vector of unknown coefficients, η is the intercept term.

A useful way of looking at the above specifications is in terms of a binary model since each individual in the sample can be thought of as contributing a maximum of five observations to the likelihood function. The binary variable will take the value of one if the individual drops out during the year and zero otherwise. To be more specific, let Ti be the recorded duration in years. Define a set of indicator variables, c and f such that, f = 1 if the individual drops out of the program in the first four years, and f= 0 otherwise; c = 1 if the individual successfully completes the program, and c= 0 if he/she drops out in the final year. Then the log-likelihood contribution by individual i with a recorded duration of T_i is given by

$$ln\, L_i = f_i\, ln[h_{1i}(T_i)] + (1-f_i)(1-c_i)\, ln[h_{2i}] - \left(\sum_{t=2}^{\min(T_i,4)} ln[1-h_{1i}(t-1)]\right) - c_i\, ln[1-h_{2i}] \quad (5)$$

---

[4] See Narendranathan and Stewart (1993a, 1993b) for a model of unemployment duration in discrete time.

[5] Because of possible endogeneity, no distinction is made between those who had an intercalated year to

Note that since there is no unobserved heterogeneity in the above model, the likelihood function (5) factors into two parts, where the parameters of the first and the second hazards can be estimated separately.[6]

The main advantage of working within the binary variable framework is that we are able to relax the extreme value assumption and use standard models such as probit and logit. Unlike the extreme value distribution probit and logit distributions are symmetric with respect to their means. Although probit and logit are symmetric distributions, they do differ in the tail behaviour.

Unobserved heterogeneity

It is well known that failure to control for any unobserved individual specific effects that may affect the hazard function will result in misleading inference due to inconsistent parameter estimators (Lancaster, 1990). The previous model can be extended for this purpose by including a random error term along with the vector of characteristics **x**. This requires an assumption regarding the distribution of this unobservable individual-specific error term. In none of the models we estimated could we find any evidence of unobservables. The models with unobservables always converged to the same point as the models without unobservables. Therefore, we report results only from models without unobservables.

4. Data and Variables

The data set is based on administrative data from the anonymised individual Universities Student Records (USR) for the full populations of undergraduate students leaving

---

complete a science degree and those who did not. This is discussed in more detail later in the main text.

[6] The above model also can be thought of in a competing-risks framework where the two risks faced by the individuals are completion and dropping out. Since the degree program lasts for a minimum of five years, the completion specific hazard has to be set equal to zero for the periods up to the 5 years. This model then collapses to the one specified above.

university in the UK in one of the academic years 1985-1993. The full data-set contains information on about 720,000 students: about 90,000 per cohort. From information on each of these ‘leaving cohorts’, we have generated a data-set comprising all those full-time students who entered university at the start of the academic year 1985 or 1986 to study for a medical degree and who had either completed their course by the end of July 1993, or left the medical degree programme prior to completion.

The reason for the choice of starting years 1985 and 1986 is based on data considerations. Data availability restricts us to cohorts leaving university no later than 1993. In general, a medical degree in the UK takes about five years to complete and the analysis is therefore conducted on students who enrolled for a five year degree programme. Normally, the first two years are classified as pre-clinical and the latter three as clinical parts of the degree. At the end of the first two years, those who perform well are given a chance to take an extra year in order to complete a Bachelor of Science degree before continuing with their medical degree. These students will then have taken 6 years to complete their original degree. Students are also allowed to retake any failed exams during their course of studies. In order to proceed to the next year of the degree program, the student is required to pass the exams (either in the first attempt or after the re-sits). Students who completed their medical degree programme in 1993 after 5 years of study would have first enrolled in 1988. However, if we study only these students, we will fail to observe students taking more than 5 years to complete. For this reason, we prefer to consider students who enrolled no later than 1986 as this gives us a possible 7 years over which to observe their withdrawal or completion. In order to increase the size of our dataset, we also include the 1985 starting cohort. The two cohorts will have faced very similar labour market and related conditions. This becomes less true if we take additional cohorts.

In the event of non-completion, an administrative leaving date is recorded along with a university-recorded reason for the student's withdrawal. From this information, we find that degree course transfers account for around 30% of all withdrawals, with academic reasons accounting for 32% and other reasons accounting for 37%. This breakdown appears to conflict with that suggested by McManus *et al*., 1995; however, for a number of reasons, we are not satisfied that the administrative coding of date and reason for leaving are reliable. For example, it is very likely that the first indication that a student has withdrawn will be their absence from examinations or their failure to submit other work. It is not clear whether this is – or should be - coded as academic failure, and in any case practice may vary across institutions. Accordingly, we have not used the actual administrative leaving date and the reason for the withdrawal in the analyses presented in this paper.

In this paper, an individual is assumed to have successfully completed if s/he obtains a medical degree by the end of five or six years regardless of whether the individual had to re-sit some examinations. Dropout is defined as withdrawal from the medical degree programme for whatever reasons. In particular, if a student changes the degree programme half way through then, this student is deemed to have withdrawn from the medical programme. All those who dropped out of the program in years five and six are assumed to have dropped out in their final year. That is, by the end of the fifth year there is assumed to be a forced termination: a successful completion, or a dropout.

Table 1 provides definitions and summary statistics for the variables used in the analyses. The 1985 cohort consists of 3,889 students, while the 1986 cohort has 3,900 students. The unconditional non-completion rate for all students is 10.7% (and is the same for each of the two years individually). This non-completion rate compares with a value of 8.9% obtained by Smith and Naylor (2001) looking across all three and four year degree courses for students commencing a degree at the start of the academic year in 1989.

Looking at the conditional dropout rates of medical students, we see that this is indeed a rare event. In particular, as the student progresses through the programme, the aggregate figures decline quite dramatically. About 50% of those who do not complete their medical degree leave prior to the start of their 2nd year. This finding is very similar to that found by Smith and Naylor (2001), where across all students, the equivalent figure was 55%, and is similar to that found in the U.S. (Porter, 1990) across all students. After the first year, the conditional rate of withdrawal declines. In particular, the conditional dropout rate in the final year of the program is only 0.37%. Given these small numbers, we have combined together both the two cohorts and also years four and five of the programme together.

Looking at the characteristics of medical students, 67% are aged 18 or less on entry and there is an approximately equal split of male and female students. Overseas fee-paying students account for only 4% of the population. Around 34% of students went to a private school, markedly higher than the 27% observed for the group of all (non-medical) students used in Smith and Naylor (2001). The average A-level (Higher) score from the best three (five) subjects was 25.4 (13.7) points (equivalent to grades of around ABB), with some 27% of the cohort obtaining the maximum 30 (15) points.[7] Some 4% of students had no prior qualification recorded and 4% already had a degree.[8] Approximately 50% of students arriving for a medical degree had either A-levels or Scottish/Irish Highers in Chemistry, Physics and Biology combination of subjects.

The social class background of students entering a medical degree programme shows that around 35% come from Social Class I (Professional) (with 13.7% actually

---

[7] Could we include a description of AL and Higher here?

coming from a background in which one of the parents or the guardian is a medical practitioner) and 38% for Social Class II (intermediate). This compares with only 19% of all students starting a 3 or 4-year degree in 1989 coming from a Professional background and 62% from either Professional or Intermediate.

5. Results

In order to achieve a satisfactory number of dropouts in each year of the programme, the models have been estimated by combining the two cohorts, with a year dummy included to account for any aggregate effect. The dependent variable takes the value of one in any one period if the individual drops out of the programme and zero otherwise. Because of the very small number of withdrawals in years five and six, the final hazard refers to years four, five, and six. We thus have one hazard specification for years 1 to 3 (see h_1 in equation (2)) and another hazard specification for the rest of the period (h_2 in equation (4)).

We first present in Table 2 the maximised log likelihood values for a set of estimated models. When the three types of model – Extreme value, logit and probit - are estimated without any covariates, the maximised values are exactly the same. The differences caused by the distributional assumptions become apparent when covariates are included in the model. The likelihood values in the second panel of Table 2 refer to a model that included the full set of covariates. Among the symmetric distributions (logit and probit), probit performs very poorly. But there is no difference between the extreme value model derived from the proportionality assumption for the underlying continuous time hazard and the logit model that does not impose this assumption. The restriction that the effects of the covariates are the same in both the hazards in the logit specification is easily

---

[8] An analysis excluding these 4% of individuals who had no qualifications did not change the results.

rejected by the data (see bottom panel of Table 2) with a $\chi^2(36)$ value of 162.3 and p-value of [0.00]. Henceforth, we shall report the results from the logit models.

The derived marginal effects on the rates of withdrawal and the corresponding p-values for the estimated logit models are reported in Table 3. These are calculated at the means of the covariates. Model 1 includes a set of university characteristics among the covariates. The results from Model 1 are presented in columns [1] and [2]. Results from substituting these university variables by a set of university dummies, called Model 2, are presented in columns [3] and [4]. Columns [1] and [3] refer to the first period hazard (h_1), and [2] and [4] to the second period hazard (h_2). As seen earlier (in Table 2), the restriction that both hazards are the same is easily rejected. We first discuss the results for Model 1 and then highlight the differences between the two models.

We did not find any significant age effects except for students older than 20 at the enrolment date. The estimated effect suggests that conditional on progressing to the final years of the programme, an individual who was more than 20 at the start of the programme has an increased probability of withdrawal at the last stages of the programme. More specifically, relative to someone under 21 at the time of enrolment, this individual is estimated to have an increased withdrawal rate of 1.6 percentage points higher, *ceteris paribus*. Males are found to be significantly more likely to withdraw relative to females, but only in the latter parts of the programme. Although UK students are less likely to withdraw in the first few years of the programme, they are found to be more likely to withdraw in years 4 and 5. Non-UK fee-paying students – a category which excludes students from the European Union - are less likely to withdraw during the first three years of their medical degree. We did not find any significant effect on withdrawal probabilities associated with the type of school attended.

There are very strong and well-determined coefficients on prior qualifications. These effects are picked up by various binary indicator variables on the type of prior qualifications as well as by the actual scores obtained for students who had taken either 'A' Level or Higher qualifications.

We first look at the effect of having completed another degree prior to starting this medical programme. One reason why this is a particularly interesting variable lies in the fact that the UK government has recently funded two postgraduate medical training centres with fast-track four-year programmes aimed at individuals who already possess biological science degrees. We find that these students are much less likely to drop out within the first three years relative to those students who have some other qualification (excluding their 'A' Level/Higher scores) *ceteris paribus*. Prior qualifications data includes information only on the individual's highest academic qualifications – not their full profile of previous attainment. Therefore, the very strong negative direct effect of 1.8 percentage points, is additional to any indirect effect associated with 'A' Level and Higher scores of students with a previous degree. However, the effect of having a previous degree is estimated to be not significant after the first three years. Because of their previous successful university experiences, students with a previous degree are hypothesised to be more able to persevere with the very strenuous and lengthy medical programme: in a sense, they have already signalled themselves as such.

With respect to performance at 'A' Level or in Highers, we note that although it is customary in UK Medical Schools to require a student to obtain three (five) 'A' Level (Higher) passes prior to entry, some students do more subjects than the required number. In order to allow for this, we have included the actual scores they obtained in their best three (five) 'A' Level (Higher) exams and also the scores from the rest if they had more than the required number. As expected, the effect from A-levels (and Highers) on the dropout rates during the entire degree programme period is estimated to be negative and declining after the first three years. An extra two points on the ‘A’ Level average (equivalent to an extra ‘A’ Level grade) reduces the drop-out probability by about 0.6 percentage points, *ceteris paribus*. A similar negative effect is also found for Highers. These negative effects are somewhat reduced if the individual also has more than the customarily required three (five) ‘A’ Level (Higher) subjects.

Looking at A-level (Higher) subjects, assuming that a strong background in Biology, Chemistry and Physics is likely to be desirable, we include three dummy variables for those students who have one, two, or three of these subjects. The base category is an individual who does not possess even one of these ‘A’ Level (Higher) scores. Relative to those with three of these subjects, student with only two of the subjects are slightly more likely to withdraw relative to those without any. However, students with only one of the favoured subjects are less likely to withdraw, *ceteris paribus*. Students with no recorded qualification or with qualifications that could not be categorised are also less likely to dropout over their entire degree course. These are in addition to those effects coming via possible ‘A’ Level and Higher scores.

The social class effects are found to be weak in general and show little effects except for a significant negative effect for individuals from Social Class category ‘Intermediate’, on the withdrawal rate in the latter parts of the program for students. Student whose parent is a medical practitioner is less likely to withdraw, although the effect is not well determined.

There are substantial effects associated with the actual medical school. Withdrawal rates are generally higher in larger medical schools, although this effect will in part be offset if the medical school has a large number of research postgraduate students. Schools in which expenditure on academic salaries is high tend to have higher withdrawal rates.

Schools in which there is a high proportion of Professors (or Research staff) while initially having a beneficial effect on the withdrawal probability in later years this effect is harmful.

We now turn to the results for Model 2 where the university characteristics are replaced by a set of university dummies. The results are reported in Columns [3] and [4] of Table 3. The estimated marginal effects are broadly similar across the two models.

6. Concluding remarks

The results show that the decision to withdraw is not random and is strongly determined by both personal characteristics and more so by prior qualifications. Students with high A-level (Higher) grades are markedly less likely to withdraw from their medical degree. Surprisingly, this effect persists over the entire period of the degree. A-level subjects are important in preparing students for their degree. In that it is better to have all science subject compared to just having 2 sciences. However, students with only one of the favoured A-levels is beneficial in withdrawal probabilities. There are few social class effects, but males are more likely to dropout. Age is important.

References

Angel, C. and Johnson, A., (2000), "Broadening access to undergraduate medical education," *British Medical Journal*, Volume 321, pp. 1136-8.

Booth, A. L. and Satchell, S. E. (1995) – 'The hazards of doing a PhD: an analysis of completion and withdrawal rates of British PhD students in the 1980s', *Journal of the Royal Statistical Society, Series A,* (Statistics in Society), Volume 158, 2, p.297-318.

Medical Workforce Standing Advisory Committee. (1997) *Planning the medical workforce: third report.* London: Department of Health.

Johnes, J. and Taylor, J., (1990). *Performance indicators in higher education*, SRHE/OUP, Oxford.

Lancaster, T., (1990) – The Econometric Analysis of Transition Data, Econometric Society Monograph, Cambridge University Press.

McManus, I. C., (1996), "Drop-out rate in medical school seems reasonable," (letter to the Editor) *British Medical Journal*, Volume 313, July, p. 173.

McManus, I.C., Richards P., and Winder, B.C., (1999), "Intercalated degrees, learning styles, and career preferences: prospective longitudinal study of UK medical students," *British Medical Journal*, Volume 319, pp. 542-6.

McManus, I.C., Richards P., Winder, B.C., Sproston, K.A. and Styles, V., (1995), "Medical school applicants from ethnic minorities: identifying if and when they are disadvantaged," *British Medical Journal*, Volume 310, pp. 496-500.

Mealli, F., Pudney, S. and Thomas, J. (1996) – 'Training duration and post-training outcomes: a duration-limited competing risks model', *The Economic Journal*, Volume 106, p. 422-33.

Narendranathan, W. and Stewart, M. B. (1993a), "How does the benefit effect vary as unemployment spells lengthen?", *Journal of Applied Econometrics*, Volume 8, p.361-81.

Narendranathan, W. and Stewart, M. B. (1993b), "Modelling the probability of leaving unemployment: competing risks models with flexible base-line hazards", *Journal of the Royal Statistical Society, Series C,* (Applied Statistics), Volume 42(1), p.63-83.

Parkhouse, J., (1996), "Intake, output, and drop out in United Kingdom medical schools," *British Medical Journal*, Volume 312, April, p. 885.

Porter, O., (1990), Undergraduate completion and persistence in four-year colleges and universities, The National Institute of Independent Colleges and Universities, Washington, D.C.

Van Ours, J. C. and Ridder, G. (2000) – 'Fast track or failure: a study of the completion rates of graduate students in economics', Centre for Economic Policy Research, Discussion Paper, No. 2363.

Smith, J.P. and Naylor, R.A., (2001), "Dropping out of university: a statistical analysis of the probability of withdrawal for UK university students," *Journal of the Royal Statistical Society, Series A,* (Statistics in Society) forthcoming.

Tinto (1975), "Dropout from higher education: a theoretical synthesis of recent research," *Review of Educational Research*, vol. 45, pp. 89-125.

Tinto (1987), Leaving college: rethinking the causes and cures of student attrition, Chicago, University of Chicago Press.

Table 1 - Definitions and Descriptive Statistics of Variables

Variable Name	Definition	Mean (standard deviation)
Individual Attributes		
Binary Age indicators	Reference category – aged 18 or less in September 1985	
Aged 19 on entry	Aged 19 in September 1985	0.157
Aged 20 on entry	Aged 20 in September 1985	0.029
Aged > 20 on entry	Aged more than 20 in September 1985	0.145
Sex - male	=1 if the student is male	0.543
Marital Status	=1 if the student is married	0.020
Nationality	=1 if the student has British Nationality	0.942
Non UK fee	=1 if the student pays non-UK fee	0.042
Type of School Attended – binary indicators	Type of school attended prior to entry into the medical program. Reference category – Grammar School which is state-funded but selection to which is typically based on ability	
Local Education Authority	=1 if local education authority school which is non-selective	0.381
Independent School	=1 if Independent School which is fee paying and also selective	0.343
College of Further Education	=1 if College of Further Education	0.080
Other	Residual category which includes Church Schools	0.062
Entry Qualifications	Type of qualifications the individual had on entry into the program.	
Already has a degree	=1 if already in possession of a degree either from a UK or an overseas institution	0.039
'A' Level scores - total	Best of three total 'A' Level scores (average among those with these qualifications). This is a university entrance-level qualification, which is typically taken at about age 18. The scores are recorded in steps of two and goes from 2 to 10. It is normal for a student to be doing three of these.	25.41(4.93)
'H' Level scores - total	Best of five total Scottish/Irish Higher scores (average among those with these qualifications). This is also a university entrance-level qualification, which is typically taken at about age 18. The scores are recorded in steps of one and goes from 1 to 3. It is normal for a student to be doing five of these.	13.71(1.88)
Top Score	=1 if the total score was the highest achievable (out of three for 'A' level subjects, and out of five for 'H' level subjects	0.272
One favoured subject	=1 if subjects taken included one from (Chemistry, Physics, Biology) among the 'A', 'H' subjects	0.051
Two favoured subjects	=1 if subjects taken included two from (Chemistry, Physics, Biology) among the 'A' 'H' subjects	0.463
Three favoured subjects	=1 if subjects taken included Chemistry, Physics, and Biology among the 'A', 'H' subjects	0.502
Other/No qualification	=1 if the individual did not have any prior qualification or had qualifications which could not be classified	0.041

Table 1 Continued

Variable Name	Definition	Mean (standard deviation)
Parental Social Class	Binary indicators for the social class of the head of household. Reference category is professional.	
Intermediate		0.383
Skilled Non-manual		0.080
Manual		0.126
Other workers		0.035
Non-workers		0.033
Father is a Doctor	=1 if the parent/guardian was a medical practitioner	0.137
University Information	Information regarding the medical faculty	
Number of undergraduates		1992
Number of postgraduates on taught degree programmes		525
Number of postgraduates on research programmes		365
Academic Salaries	Expenditure on academic salaries £000 per medical student	12.38
Other Expenditure	£000 per medical student	7.38
Base: Junior Staff		
Professors	% of staff who are professors	10.53
Senior Staff	% of staff who are senior lectures or readers	25.80
Research	% of staff who are on research scale	36.82
Number of Students (initial)		
1985 entry cohort		3889
1986 entry cohort		3900
Conditional dropout rates		
1985 entry cohort – %		
Year 1		5.32
Year 2		3.04
Year 3		1.01
Year 4		1.33
Year 5		0.37
Unconditional dropout rate %		4.67
Conditional dropout rates		
1986 entry cohort - %		
Year 1		5.54
Year 2		2.39
Year 3		1.39
Year 4		1.47
Year 5		0.37
Unconditional dropout rate %		4.67

Table 2

Maximised Likelihood Values

Model	**Maximised Log Likelihood Value**
Intercept only models[1]	
1. Discrete time hazard is Extreme Value	-3656.38
2. Discrete time hazard is Logit	-3656.38
3. Discrete time hazard is Probit	-3656.38
Models with full set of covariates including university characteristics	
1. Discrete time hazard is Extreme Value	-2705.16
2. Discrete time hazard is Logit	-2705.89
3. Discrete time hazard is Probit	-2758.41
Logit Model with the restriction $\beta = \alpha$ (36 restrictions)	-2787.03

Notes:
1. Model has parameters δ and η.

Table 3 – Derived Marginal Effects [p-values] on the Conditional Exit rate of Withdrawal from the Medical Degree Programme*

Logit Hazard Models

Variable	With university variables - MODEL 1		With university dummies - MODEL 2	
	Years 1 to 3 of the Programme (h_1)	Years 4 to 5 of the Programme (h_2)	Years 1 to 3 of the Programme (h_1)	Years 4 to 5 of the Programme (h_2)
Intercept	3.667 [0.00]**	-3.122 [0.04]**	1.775 [0.00]**	-4.654 [0.00]**
Year 2 dummy	-0.453 [0.00]**		-0.441 [0.00]**	
Year 3 dummy	-0.677 [0.00]**		-1.107 [0.00]**	
1986 Year Dummy	-0.301 [0.08]*	-0.037 [0.92]	0.210 [0.04]**	0.272 [0.23]
Aged 19 on entry	-0.042 [0.81]	0.494 [0.22]	-0.118 [0.50]	0.527 [0.19]
Aged 20 on entry	-0.332 [0.39]	0.447 [0.54]	-0.509 [0.21]	0.601 [0.40]
Aged > 20 on entry	-0.286 [0.13]	1.646 [0.00]**	-0.266 [0.16]	1.615 [0.00]
Sex - male	0.041 [0.68]	0.754 [0.00]**	0.027 [0.79]	0.736 [0.00]**
Married	0.353 [0.26]	-0.192 [0.00]	0.389 [0.22]	-0.256 [0.65]
British National	-0.482 [0.05]**	1.353 [0.02]**	-0.479 [0.05]**	1.595 [0.01]**
Non UK fee student	-0.837 [0.01]**	0.775 [0.12]	-0.703 [0.03]**	0.912 [0.06]*
Type of School Attended – binary indicators [base Local Education Authority]				
Grammar School	-0.141 [0.45]	-0.400 [0.30]	-0.080 [0.66]	0.075 [0.84]
Independent School	0.190 [0.13]	-0.105 [0.71]	0.158 [0.21]	-0.201 [0.49]
College of Further Education	0.254 [0.20]	-0.715 [0.23]	0.257 [0.20]	-0.702 [0.24]
Other (Church Schools and other)	-0.203 [0.51]	0.403 [0.45]	-0.146 [0.64]	0.466 [0.39]
Entry Qualifications				
Already has a degree	-1.802 [0.00]**	-0.228 [0.60]	-1.883 [0.00]**	-0.256 [0.56]
'A', 'AS' Level scores – best of 3 (max30)	-0.264 [0.00]**	-0.174 [0.00]**	-0.263 [0.00]**	-0.167 [0.00]**
'H' Level scores - best of 5 (max 15)	-0.380 [0.00]**	-0.278 [0.00]**	-0.399 [0.00]**	-0.308 [0.00]**
Top Score – binary indicator	-0.035 [0.84]	0.051 [0.88]	0.028 [0.87]	0.228 [0.52]
One favoured subject (Chem/Phy/Bio)	-0.074 [0.78]	-0.164 [0.83]	-0.063 [0.82]	-0.036 [0.96]
Two favoured subjects (Chem/Phy/Bio)	0.567 [0.02]**	1.342 [0.02]**	0.496 [0.05]**	1.560 [0.01]**
Three favoured subjects (Chem/Phy/Bio)	0.181 [0.72]	0.585 [0.32]	0.101 [0.69]	0.808 [0.18]
'A'/'AS' Level scores – excl. the best 3	0.117 [0.00]**	0.087 [0.00]**	0.120 [0.00]**	0.090 [0.00]**
'H' Level scores – excl. best 5	0.031 [0.71]	0.113 [0.59]	0.018 [0.83]	0.171 [0.41]
Other qualification	-3.844 [0.00]**	-2.097 [0.00]**	-4.072 [0.00]**	-1.781 [0.02]**

Table 3– Continued

Variable	With university variables - MODEL 1		With university dummies (18) - MODEL 2	
	Years 1 to 3 of the Programme (h_1)	**Years 4 to 5 of the Programme (h_2)**	**Years 1 to 3 of the Programme (h_1)**	**Years 4 to 5 of the Programme (h_2)**
Parental Social Class – binary indicators [Base is Professional]				
Intermediate	-0.057 [0.68]	-0.614 [0.03]**	-0.089 [0.52]	-0.608 [0.04]**
Skilled Non-manual	0.205 [0.30]	-0.370 [0.41]	0.183 [0.35]	-0.398 [0.38]
Manual	0.019 [0.92]	-0.417 [0.29]	0.017 [0.93]	-0.407 [0.30]
Other workers	-0.178 [0.52]	-0.750 [0.28]	-0.188 0.50]	-0.660 [0.34]
Non-workers	0.456 [0.16]	-0.091 [0.87]	-0.431 [0.19]	-0.184 [0.74]
Father/Guardian is a Doctor	-0.158 [0.37]	-0.320 [0.40]	-0.175 [0.32]	-0.392 [0.30]
Medical Faculty Information				
Number of undergraduates (10^{-2})	-0.006 [0.90]	0.003 [0.02]**		
No. p/g on taught degree programmes (10^{-2})	0.023 [0.85]	0.008 [0.00]**		
No. of p/g on research programmes (10^{-2})	-0.011 [0.98]	-0.026 [0.00]**		
Expenditure on Salaries per medical student (£000)	0.043 [0.00]	-0.000 [0.99]		
Expenditure on grants and other per medical student (£000)	-0.065 [0.06]*	-0.007 [0.92]		
Professors (% of all staff)	-0.121 [0.00]**	0.174 [0.01]**		
Senior Staff (% of all staff)	-0.003 [0.98]	-0.094 [0.01]**		
Research (% of all staff)	-0.003 [0.70]	0.040 [0.01]**		
Maximised log likelihood value	-2705.89		-2696.08	
P-value for the additional university variables	[0.00]		[0.00]	
Number of observations	7789		7789	

Notes: * Since the probabilities are small, the marginal effects are reported in terms of percentage effects on the conditional withdrawal rates.

Table 4 - Marginal Effects x 100 [p-value]

Entry Qualifications	Simple Logit	MODEL 1		MODEL 3		MODEL 4	
		Years 1 to 3 of the Programme (h_1)	**Years 4 to 5 of the Programme (h_2)**	**Years 1 to 2 of the Programme (h_1)**	**Years 3 to 5 of the Programme (h_2)**	**Year 1 of the Programme (h_1)**	**Years 2 to 5 of the Programme (h_2)**
Already has a degree	-6.086 [0.00]**	-1.802 [0.00]**	-0.228 [0.60]	-2.801 [0.00]**	-0.446 [0.39]	-4.224 [0.00]**	-2.173 [0.00]**
'A', 'AS' Level scores – best of 3 (max30)	-1.272 [0.00]**	-0.264 [0.00]**	-0.174 [0.00]**	-0.413 [0.00]**	-0.209 [0.00]**	-0.612 [0.00]**	-0.516 [0.00]**
'H' Level scores - best of 5 (max 15)	-1.845 [0.00]**	-0.380 [0.00]**	-0.278 [0.00]**	-0.588 [0.00]**	-0.321 [0.00]**	-0.911 [0.00]**	-0.747 [0.00]**
'A'/'AS' Level scores – excl. the best 3	0.630 [0.00]**	0.117 [0.00]**	0.087 [0.00]**	0.165 [0.00]**	0.136 [0.00]**	0.240 [0.00]**	0.262 [0.00]**
'H' Level scores – excl. best 3	0.253 [0.51]	0.031 [0.71]	0.113 [0.59]	0.023 [0.86]	0.183 [0.41]	0.373 [0.11]	-0.201 [0.43]
Other qualification	-19.546 [0.00]**	-3.844 [0.00]**	-2.097 [0.00]**	-5.841 [0.00]**	-2.987 [0.00]**	-8.918 [0.00]**	-7.535 [0.00]**

Notes:
1. All Models include University Characteristics and also the other controls as before.
2. The marginal effects on the conditional probability of exiting in period 2 (h_2) is calculated conditional on survival up to that point.

Table 5 - Predicted probabilities (x 100) for some individuals

(Predicted probability of dropping out in year 1x 100)

Characteristic changed for the reference individual	Simple Logit	Model 1 [1,2,3] [4,5]	Model 3 [1,2] [3,4,5]	Model 4 [1] [2,3,4,5]
1. Reference individual[1]	2.54	1.28 (0.42)	2.90 (1.19)	2.91 (1.02)
2. Entered in 1986	1.82	1.07 (0.32)	2.32 (0.87)	2.67 (0.94)
3. Has AAB	1.56	0.85 (0.26)	1.91 (0.71)	1.88 (0.63)
4. Has AAA (topscore)	0.89	0.58 (0.16)	1.31 (0.44)	1.14 (0.36)
5. Has BBB	4.10	1.97 (0.68)	4.46 (1.98)	4.51 (1.68)
6. Woman	2.22	1.05 (0.41)	1.31 (1.93)	2.43 (1.07)
7. Only has two of the preferred subjects	3.75	2.06 (0.60)	4.40 (1.61)	4.37 (1.72)

Notes: 1. The reference individual is a man, British national, has 3 'A'Levels (in Phyics, Chemistry & Biology), with a total score of 26 (ABB) and comes from social class 'Intermediate', and entered the medical school with average characteristics, in 1985.

2. [1,2,3][4,5] means that the first period hazard is defined over the first three years and the completion/non-completion hazard is defined over years 4 and 5.