

A hazard model of the probability of medical school drop-out in the UK

Wiji Arulampalam

University of Warwick, Coventry, UK, and Institute for the Study of Labour, Bonn, Germany

and Robin A. Naylor and Jeremy P. Smith

University of Warwick, Coventry, UK

[Received May 2002. Revised June 2003]

Summary. From individual level longitudinal data for two entire cohorts of medical students in UK universities, we use multilevel models to analyse the probability that an individual student will drop out of medical school. We find that academic preparedness—both in terms of previous subjects studied and levels of attainment therein—is the major influence on withdrawal by medical students. Additionally, males and more mature students are more likely to withdraw than females or younger students respectively. We find evidence that the factors influencing the decision to transfer course differ from those affecting the decision to drop out for other reasons.

Keywords: Competing risks; Limited duration model; Medical students; Multilevel models; Student drop-out (non-completion) probabilities; Survival analysis

1. Introduction

The issue of the determinants of medical student drop-out probabilities is important and topical in both the UK and beyond for a variety of reasons. First, there is serious and growing concern in the UK and elsewhere regarding a shortage in the domestic supply of medical doctors. The third report of the UK's Medical Workforce Standing Advisory Committee observed that there will be a deficit of doctors in Europe early in this century (Medical Workforce Standing Advisory Committee (1997), page 30). This has led to Government-supported enquiries into the causes and potential cures for this problem. The most obvious policy initiative is to train more doctors. The Medical Workforce Standing Advisory Committee (1997) concluded that a substantial increase in medical school intake was required to meet the future workforce needs of the UK National Health Service. In response, the Government allocated 1129 new medical places, in a three-stage process beginning in 1999. The subsequent National Health Service 'Plan 2000' announced the Government's target of a further 1000 medical places, with the expectation that the majority of the additional student intake will be in place by 2006. These changes imply that the total medical school intake will have almost doubled over a 10-year period from 1997. As has been highlighted in policy discussions (Medical Workforce Standing Advisory Committee, 1997), an expansion in the numbers of medical students begs questions regarding both the quality and the retention of students. This paper attempts to inform our understanding of the latter issue.

Address for correspondence: Robin A. Naylor, Department of Economics, University of Warwick, Coventry, CV4 7AL, UK.
E-mail: robin.naylor@warwick.ac.uk

A second reason for examining medical student drop-out behaviour relates to the UK debate on the desirability of 'widening' access into higher education, in general, and into medical schools in particular. There has been a lively and high profile debate in the UK concerning the extent of accessibility of medical schools to students regardless of their social or school background (see McManus (1998) and the related discussion). This has led to explicit recommendations to broaden access to undergraduate medical education (Angel and Johnson, 2000). Predicting the likely effect of such policies on retention and progression is clearly an important issue. A third reason for analysing withdrawal by medical students is related to the debate concerning the extent to which previous educational qualifications affect the performance and progression of medical students (see, for example, McManus *et al.* (1999)). Additionally, Arulampalam *et al.* (2003) analyse the determinants of first-year drop-outs over the period 1980–1992. An issue here concerns whether it should be compulsory for medical students to have studied science subjects before enrolment.

The rest of this paper is organized as follows. Section 2 presents a discussion of the institutional and policy contexts, which provide the backdrop to our analysis of data on UK medical students. Section 3 describes the data set. Section 4 presents the econometric model together with a discussion of relevant issues regarding our model estimation procedure. We note that our study is the first in the literature to apply a hazard model to the analysis of withdrawals by undergraduate students. We also observe that no previous analysis of withdrawals by medical students in the UK has exploited data on full cohorts of medical students. In Section 5, we present estimates of the determinants of drop-out probabilities based on a single-risk hazard model. The basic drop-out probability model that is presented in Sections 4 and 5 is then extended in Section 6 to a more detailed analysis of the underlying reasons for withdrawal through the application of a 'competing risks' framework. Finally, Section 7 closes the paper with conclusions and further remarks.

2. Institutional context and public policy

Labour economics focuses chiefly on the analysis of decentralized 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. In important respects, however, the labour market for medical practitioners in the UK is better described by a planning model than by a model of a purely decentralized market. Primarily, this is because of the nature of the necessary regulation in the market. In the UK, doctors are trained within medical schools that are 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. This is a very different approach from that which has recently been recommended from an analysis of the shortage of post-graduate 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 Medical Workforce Standing Advisory Committee (1997) 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 time lags between forecasting future supply and demand and the subsequent 'production' of the medical workforce. Uncertainties in demographic trends

and in the evolution of medical technologies compound the difficulties that are associated with forecasting and planning.

The Medical Workforce Standing Advisory Committee (1997) proposed various 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 in the intake of medical students (about 1000 per annum) together with policies to ensure 'minimized levels of wastage from such courses, thereby increasing the proportion of entrants who qualify as doctors'. Of course, a significant incentive to reduce wastage rates lies in the high cost that is associated with medical training. There is no consensus figure on the full cost of a medical training—partly because the costs vary between institutions and are borne by a variety of parties—but a figure of around £50 000 per annum has often been quoted. One element of the proposed package of measures to address the issue of minimizing wastage concerns changes in the selection procedures of candidates for medical schools 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 debate.

As the Medical Workforce Standing Advisory Committee (1997) emphasized, the development of strategies to minimize the drop-out rate of medical students should be conducted in the light of an 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 Medical Workforce Standing Advisory Committee (1997) acknowledged that there is significant uncertainty regarding estimates of qualification rates for medical schools. Estimates are typically based on the difference between 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 drop-out rates, which fluctuated (with no systematic trend) between 8% and 14% during the period from 1986–1987 to 1991–1992 (see Medical Workforce Standing Advisory Committee (1997), page 65). So severe is the information problem regarding the drop-out rate that the Medical Workforce Standing Advisory Committee (1997) acknowledged that all its analysis and recommendations rested on an estimated wastage rate of students which is itself based on 'some anecdotal evidence that drop-out is falling' (page 25). One of the eight recommendations of the report called for more information and research into wastage rates from medical schools, *inter alia*.

In the absence of reliable data on drop-out rates from medical schools, different studies have produced widely varying estimates of the average national rate. Parkhouse (1996) used Universities' Statistical Record and Higher Education Funding Council for England data on the intakes of medical schools and numbers qualifying 5 years later and estimated an average drop-out rate in the UK of between 11.7% and 14.1%. McManus (1996) disputed these figures and cited 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.

Previous data and accompanying analyses have been based, typically, either on aggregated (medical school level) official data from the Universities' Statistical Record or Higher Education Funding Council for England or on follow-up surveys of particular subsamples of medical students. This mirrors the situation regarding the analysis of all UK university students across all subject areas. Very recently, however, (anonymized) individual student level administrative data for full population cohorts of students have become available to researchers for the period 1972–1993. These data contain rich information not only on the academic characteristics of

students in UK universities (their courses, institutional affiliation, performance and reasons for leaving, *inter alia*) but also on their personal, social and prior educational characteristics. These data offer the prospect of much more precise estimates of drop-out rates and of detailed analytical investigation of the factors that are associated with drop-out behaviour. This is of clear interest and relevance to the issue of examining the effect of changing selection procedures on medical students' drop-out probabilities (see, for example, Angel and Johnson (2000)). Given the importance of the issue to public policy, in the current paper we focus exclusively on a detailed analysis of the determinants of the probability of withdrawal of UK medical students.

Individual UK student level data have previously been exploited to analyse academic performance and drop-out probabilities of students on 3- or 4-year degree programmes (Smith and Naylor, 2001a, b). These studies have omitted medical students for a variety of reasons, including the fact that medical courses are significantly longer than standard undergraduate programmes and the fact that medical students' characteristics are typically rather different from those of other students. The average A-level attainment, for example, is much higher for medical students, whose average A-level points score was 26 for the cohorts that are considered in this paper, compared with a score of less than 22 points for contemporaneous non-medical students—a difference of more than two A-level grades. Accordingly, it is appropriate to model the behaviour of medical students separately from that of other students.

3. The data

All higher education institutions—including all medical schools—in the UK are required annually to deposit comprehensive longitudinal individual student records data with a central government agency. Our data set is based on these administrative data from the anonymized individual Universities' Statistical Record for the full populations of undergraduate students leaving university in the UK in one of the academic years 1985–1993. 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 had left their medical degree programme before completion.

The reason for the choice of starting years 1985 and 1986 is based on considerations of the data. The availability of data restricts us to cohorts leaving university no later than 1993. In general, a medical degree in the UK takes 5 years to complete and the analysis is therefore conducted on students who enrolled for a 5-year degree programme. Traditionally, the first 2 years have been classified as preclinical and the last 3 years as clinical parts of the degree. At the end of the second year of the programme those who perform well are given the opportunity to take an extra year to complete a Bachelor of Science degree before continuing with their medical degree. If they are successful, these students will then have taken a minimum of 6 years to complete their original degree. Students are also allowed to retake any failed examinations during their course of studies. To proceed to the next year of the degree programme, the student is required to pass the examinations (either at the first attempt or after resitting them). 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 shall fail to observe students taking more than 5 years to complete the course. For this reason, we prefer to consider students who enrolled no later than 1986 as this gives us a minimum of 7 years over which to observe their withdrawal or completion. To increase the size of our data set, we also include the cohort starting in 1985. The two cohorts will have faced very similar labour market and related conditions. This becomes less true if we take additional earlier 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, among those who registered for a medical degree, course transfers account for around 26% of all withdrawals, with academic reasons accounting for 36% and other reasons accounting for 32%. From the point of view of the medical profession, what matters is the simple dichotomy between completion and withdrawal. In our basic model, this distinction forms the basis for our dichotomous dependent variable. In Section 6, we also exploit information on the reason for withdrawal within a competing risks framework.

In this paper, an individual is assumed to have successfully completed if she or he obtains a medical degree by the end of 5–7 years (and 8 years for the 1985 cohort) regardless of whether the individual had to resit some examinations. Drop-out is defined as withdrawal from the medical degree programme for whatever reason. All those who are observed to have dropped out of the programme after five programme years are assumed to have dropped out in the final year of the programme. Thus, the standard length of the medical programme, which is 5 years, is used as the duration time in our analysis rather than the actual calendar time taken for completion. Hence, by the end of the fifth year there is assumed to be a forced termination: a successful completion, or a drop-out.

The second and third columns of Table 1 provide some information about drop-out rates. The 1985 cohort consists of 3889 students, whereas the 1986 cohort has 3900 students. The unconditional non-completion rate for all students who started a medical degree in either 1985 or 1986 is around 11%. From these two cohorts of medical students, 8.1% actually dropped out from university altogether (with 2.8% dropping out of a medical degree because of a change in course). This non-completion rate of 8.1% compares with a value of 8.9% that was obtained by Smith and Naylor (2001b) looking at all 3- and 4-year degree courses for students beginning a degree at the start of the academic year in 1989.

Looking at the conditional drop-out rates of medical students, we see that, as the student progresses through the programme, the conditional drop-out rate declines quite dramatically. About 50% of those who do not complete their medical degree leave before the start of their second year. This finding is very similar to that of Smith and Naylor (2001b), where across all students the equivalent figure was 55%, and is similar to that of Porter (1990) across all students in the USA. The conditional drop-out rate in the final year of the programme is only 0.37%.

Table 1

	<i>Overall drop-out for the following years:</i>		<i>Results for the following reasons for drop-out:</i>	
	<i>1985</i>	<i>1986</i>	<i>Course change 1985 + 1986</i>	<i>Other, 1985 + 1986</i>
Number of students (initial)	3889	3900	7789	7789
Conditional drop-out rate (%)				
Year 1	5.32	5.54	1.85	3.58
Year 2	3.04	2.39	0.79	1.82
Year 3	1.01	1.39	0.20	0.92
Year 4	1.33	1.47	0.00	1.40
Year 5	0.37	0.37	0.00	0.36
Unconditional drop-out rate (%)	10.67	10.74	2.77	8.09

Table 2. Definitions and descriptive statistics of variables†

<i>Variable name and categories</i>	<i>Definitions</i>	<i>Means for the following samples:</i>		
		<i>Overall sample</i>	<i>Drop-out sample</i>	<i>Completion sample</i>
Entry age (binary indicators)				
Aged ≤ 18	Aged 18 years or less in September of entry year	0.619	0.573	0.625
Aged 19–21	Aged 19–21 years	0.320	0.311	0.321
Aged > 21	Aged more than 21 years	0.061	0.116	0.054
Sex—male	1 if the student is male	0.543	0.586	0.538
Nationality	1 if the student has British nationality	0.942	0.926	0.944
Non-UK fee	1 if the student pays non-UK fee	0.042	0.058	0.040
Accommodation (binary indicators)	Accommodation in the first year at university			
Living at home	Living at home	0.110	0.118	0.108
Living on campus	Living at university campus accommodation	0.773	0.739	0.778
Other	Not living at home or university campus	0.117	0.143	0.114
School attended (binary indicators)	Type of school attended before entry to the medical programme			
Local education authority school	School is non-selective	0.381	0.382	0.381
Grammar school	State funded, but selection is typically based on ability	0.135	0.101	0.139
Independent school	Fee-paying school and is also selective	0.343	0.333	0.344
College of further education	College of further education	0.080	0.076	0.080
Other	Residual category which includes Church schools	0.061	0.108	0.056
Main entry qualifications	Type of qualifications the individual had on entry to the programme			
A-level scores— best-of-3 total	Best of 3 total A-level scores	26.2 (3.9)	25.4 (4.3)	26.3 (3.9)
Highers scores— best-of-5 total	Best of 5 total Scottish or Irish Higher scores (average among those with these qualifications)	13.7 (1.9)	13.5 (1.9)	13.7 (1.9)
Percentage of students having more than 3 A-levels		42.9	61.4	40.7
Other A-level scores	Excludes the best 3 A-levels	8.9 (4.9)	13.3 (6.1)	8.1 (4.2)
Percentage of students having more than 5 Highers		7.2	7.7	7.1
Other Highers scores	Excludes the best 5 Highers scores	2.8 (1.3)	2.7 (1.3)	2.8 (1.3)
Top score	1 if the total score was the highest achievable (out of 3 for A-level subjects, and out of 5 for Higher level subjects)	0.272	0.139	0.288
1 favoured subject	1 if subjects taken included 1 from chemistry, physics and biology among the A-level or Higher subjects	0.013	0.016	0.012

(continued)

Table 2 (continued)

Variable name and categories	Definitions	Means for the following samples:		
		Overall sample	Drop-out sample	Completion sample
2 favoured subjects	1 if subjects taken included 2 from chemistry, physics and biology among the A-level or Higher subjects	0.439	0.474	0.435
3 favoured subjects	1 if subjects taken included chemistry, physics and biology among the A-level or Higher subjects	0.503	0.416	0.513
Other entry qualifications (binary indicators)				
A-levels or Highers only	Students who have only A-levels or Scottish Highers	0.920	0.868	0.925
UK university qualifications	Mainly students who already have a degree	0.039	0.056	0.038
Other	Students with no qualification, Business and Technology Education Council or other UK or foreign qualification	0.041	0.076	0.037
Parental social class (binary indicators)	Social class of the head of household			
Professional	Social class I	0.343	0.331	0.345
Intermediate	Social class II	0.383	0.342	0.388
Skilled non-manual	Social class III non-manual	0.080	0.094	0.078
Other	Social class III manual, social class IV and social class V: manual or other workers including non-workers	0.194	0.233	0.189
Father is a doctor	Parent or guardian was a medical practitioner	0.137	0.126	0.139
Number of students		7789	834	6955

†Standard deviations are given in parentheses.

In the UK, the usual age of entry into a university medical school is 18 years. Looking at Table 2 we see that 62% of our sample members are aged 18 years or less on entry, with a slightly lower proportion found among the drop-outs compared with those who successfully complete the course. Although our sample contains an approximately equal split of male and female students, males are found to predominate among the drop-outs.

In the two cohorts that are used in the analysis, 94% of the students were of UK nationality. The fees that are charged by universities depend on the nationality as well as on some residency conditions. In general, UK students are liable for a UK fee, which is much smaller than the overseas student's fee. In any case, for the cohorts that we are analysing in the present paper, all fees facing a UK student were paid for by the student's local education authority. A European Union student would be liable to pay the UK level of fee. In the academic year 2002–2003, this was just over £1100 per annum. The overseas fee varies across universities, but in 2002–2003 was around £9000 per annum for the two preclinical years and £17000 per annum for the three clinical years. Overseas fee paying students account for only 4% of the population, but for 6% among the drop-outs.

Before entering university, most UK students study in secondary school towards qualifications, the level of performance in which will determine the success of their applications to

higher education. Broadly, we can distinguish between two types of school: those which are in the private sector (henceforth, 'independent' schools) and those that are in the broadly defined state sector. The latter consists of various subcategories of school, including local education authority comprehensive schools, grammar schools (to which admission is selective and subject to educational tests) and colleges of further education. In the school population of the UK as a whole, about 7% of pupils attend an independent school. In contrast, around 34% of medical students went to an independent school. This is markedly higher even than the 27% that is observed for the group of all (non-medical) UK university students.

The pre-university secondary school qualifications that form the basis for offers of places at medical schools are, typically, A-levels for school pupils from England, Wales and Northern Ireland, and 'Highers' for school pupils from Scotland. A-levels are classified as A–E. These grades can be converted into a points score: A \equiv 10 points, B \equiv 8 points, C \equiv 6 points, D \equiv 4 points and E \equiv 2 points. Highers are classified as A–C: A \equiv 3 points, B \equiv 2 points and C \equiv 1 point. Offers of university places are typically conditional on the candidate's performance in their best three A-levels (or best five Highers). In addition, some medical schools interview candidates before making an offer of a place on the course.

The average A-level or Higher score from the best three or five subjects respectively was 26.2 or 13.7 points, with about 27% of the cohort having the maximum 30 or 15 points respectively. Approximately 50% of students arriving for a medical degree had either A-levels or Highers in chemistry, physics and biology. Not all medical students enter with the standard A-level or Higher qualifications. About 4% of students had no prior qualifications recorded or only had a Business and Technology Education Council or similar qualification. In addition, another 4% already had a prior qualification that had been obtained from a university in the UK.

The social class background of students entering a medical degree programme shows that around 34% come from social class I (professional) (with 39% of these actually coming from a background in which one of the parents or the guardian is a medical practitioner) and 38% from social class II (intermediate professions). This compares with only 19% of all students starting a 3- or 4-year degree coming from a professional background and 43% from an intermediate professional background.

Summarizing with respect to the characteristics of those among the drop-out sample, compared with those who successfully complete the degree programme, there is a slightly larger proportion of individuals

- (a) who are aged more than 19 years,
- (b) who are men,
- (c) who are paying non-UK fees,
- (d) who have an average A-level score which is slightly lower than ABB and
- (e) who have more than the standard three A-levels.

4. Econometric model

In the current paper, our objective is to model the probability that an individual student will drop out of a medical degree programme during some small time interval, conditionally on not having dropped out up to that point. As indicated above, the underlying variable is the time that is spent on the programme rather than calendar time. Unlike in conventional duration (survival) models, two specific characteristics of the programme need to be addressed. First, the duration of the programme is limited to 5 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 programme cannot be com-

pleted before the end of 5 years, i.e. the probability of successfully completing the programme during the first 5 years is 0. We are not aware of any previous study that applies a hazard model to the analysis of the undergraduate drop-out probability. Mealli *et al.* (1996) considered a duration-limited competing risks model in the context of youth training programmes. Booth and Satchell (1995) and van Ours and Ridder (2000) looked at a related model for completion rates for doctoral courses.

4.1. Cox’s proportional hazards model

We start with Cox’s proportional hazards model. Given the above characterization of the programme of studies, the hazard for individual i , $h_i(t)$, is parameterized as

$$h_i(t) = \lambda(t) \exp(\mathbf{x}'_i \beta) \quad (t < 5 \text{ years}) \tag{1}$$

where $\lambda(t)$ is the base-line hazard at time t , \mathbf{x}_i is the vector of characteristics for individual i (excluding the intercept term) and β is the corresponding vector of unknown coefficients. As discussed earlier, because of possible measurement errors in the recording of the date of drop-out, for the purpose of the analysis presented here, the duration information has been recoded in terms of whole years completed. 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 programme at time $t - 1$, is given by

$$\begin{aligned} q_{1i}(t | \mathbf{x}_i) &= \text{Prob}(T_i < t | t - 1 \leq T_i) = 1 - \exp\left\{-\int_{t-1}^t h_i(\tau) d\tau\right\} \\ &= 1 - \exp\left\{-\int_{t-1}^t \lambda(\tau) \exp(\mathbf{x}'_i \beta) d\tau\right\} \\ &= 1 - \exp[-\exp\{\mathbf{x}'_i \beta + \delta(t)\}] \end{aligned} \tag{2}$$

where

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

We thus have an extreme value form for the drop-out probability model in discrete time. See Narendranathan and Stewart (1993a, b) for a model of unemployment duration in discrete time.

As seen earlier, the degree programme finishes at the end of year 5. There is thus a forced termination at this point. Because of possible endogeneity, we do not make any distinction between those who had an intercalated year to complete a science degree and those who did not. To account for the forced termination, we assume that these limit point probabilities take the same extreme value form as before, but with a different set of coefficients. This is specified as

$$\begin{aligned} q_{2i}(t = 5 | \mathbf{x}_i) &= \text{Prob}(\text{dropping out in year 5} | \text{survival up to year 5}, \mathbf{x}_i) \\ &= 1 - \exp\{-\exp(\mathbf{x}'_i \alpha + \eta)\} \end{aligned} \tag{4}$$

where α is the vector of unknown coefficients and η is the intercept term.

Let T_i be the recorded duration in years for individual i . Then the likelihood contribution by individual i with a recorded duration of $T_i < 5$ is given by

$$L_i = q_{1i}(T_i) \prod_{t=2}^{T_i} \{1 - q_{1i}(t - 1)\}. \tag{5}$$

The likelihood contribution by individual i who proceeds to year 5 is given by

$$L_i = \left[(1 - q_{2i}) \prod_{t=2}^5 \{1 - q_{1i}(t - 1)\} \right]^{1-d_i} \left[q_{2i} \prod_{t=2}^5 \{1 - q_{1i}(t - 1)\} \right]^{d_i} \tag{6}$$

where d is a binary variable which takes the value of 1 if the individual drops out during the final year of the programme and is equal to 0 otherwise. The likelihood function, which will be made up of terms given by equations (5) and (6), will factor into two parts and thus will facilitate a separate estimation of the parameters of the two exit probabilities. The above model also can be thought of in a competing risks framework where the two risks that are faced by the individuals are completion and dropping out. Since the degree programme lasts for a minimum of 5 years, the completion specific hazard must be set equal to 0 for the periods up to the 5 years. This model then collapses to the model specified above.

A useful way of looking at the above specification is in terms of a binary response model. In the estimation of the first exit probability q_1 , the dependent variable can be thought of as taking the value of 0 in the years before exit and a value of 1 in the year of exit before year 5. Each individual will make a maximum of four contributions to the estimation here. An individual who proceeds to the final year will have a binary sequence, which is entirely made up of 0s. The estimation of the parameters of the second hazard is then carried out by using only the subsample of individuals who have proceeded to the final year of the programme. This again can be thought of as a simple binary response model where the dependent variable is d , as defined above.

The main advantage of working within the binary variable framework is that we can then relax the extreme value assumption and use popular models such as the probit and logit that imply a non-proportional hazard in the underlying *continuous time* framework. For example, in the case of a logit model, the conditional probabilities q_{1i} and q_{2i} will be specified as

$$q_{1i}(t) = \frac{\exp\{\mathbf{x}'_i \boldsymbol{\beta} + \delta(t)\}}{1 + \exp\{\mathbf{x}'_i \boldsymbol{\beta} + \delta(t)\}}$$

and

$$q_{2i} = \frac{\exp(\mathbf{x}'_i \boldsymbol{\alpha} + \eta)}{1 + \exp(\mathbf{x}'_i \boldsymbol{\alpha} + \eta)}$$

and these conditional probability functions are then used in equations (5) and (6). Unlike the extreme value distribution, probit and logit distributions are symmetric with respect to their means, although they do differ in the tail behaviour.

4.2. Unobserved heterogeneity

It is well known that a 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 \mathbf{x} . This requires an assumption regarding the distribution of this unobservable individual-specific error term. In addition, we also require the assumption of independence of the unobservables and the included regressors to marginalize with respect to these unobservables. The models for the unobservables that were tried were normally distributed unobservables, normally distributed unobservables with allowance for different masses at the end points and a two-mass-point discrete distribution. In none of the models that we estimated could we find any evidence of unobservable heterogeneity. Models

which allowed for unobserved heterogeneity always converged to the same point as models without unobservables, and hence we report results only from the latter.

4.3. *Multilevel models: fixed versus random medical school effects*

One further issue that needs to be addressed is the issue of possible clustering of drop-out rates within universities given the hierarchical nature of our data set. We therefore need to allow for unobservable medical school factors to affect student drop-out probabilities. We can either treat the unobservable medical school effects as fixed or as random. The first model is known as the fixed effects (FE) model and the second as the random-effects (RE) model. These models are known as multilevel models. The FE model can be thought of as providing inferences conditionally on the medical schools in the sample. In contrast, the RE model can be thought of as providing marginal inferences that will enable us to generalize our results to the population of all medical schools. We consider the FE model as more appropriate for our analysis as the data and the sample that were analysed cover the full population of medical schools in the UK. We therefore include a set of medical school indicator variables and estimate the effects as part of the parameters of the model (model 1). However, we also provide estimates that were obtained under the assumption of an RE model to make comparisons across different specifications.

There are advantages and disadvantages to the two approaches, which should be borne in mind in the comparisons of estimates across different models. The FE model does not allow us to include characteristics of the medical schools in addition to the medical-school-specific binary indicator variables; hence, we also estimate our FE model by replacing the medical school indicator variables by a set of medical school characteristics (model 2). This problem is not present in the RE model since the medical school effects are assumed to be random drawings from a particular distribution. But, to marginalize the likelihood function with respect to the unobservables in this model, we need to assume that the covariates included and the REs are independently distributed. Although the RE model treats the unobserved medical school effects as random drawings from a distribution, we could obtain the estimates of 'shrunk' residuals (or empirical Bayes estimates) of these effects (Goldstein, 2003), which can then be compared with the estimated medical school 'FE'. Because we are dealing with very large numbers of observations per university, we expect to see that the estimated medical school effects will be very similar across these models. In this specification, we assume that the unobservable medical-school-specific RE are drawn from a normal distribution. All the models were estimated by using STATA7 (Stata Corporation, 2001) and `gllamm` written by Rabe-Hesketh *et al.* (2001).

4.4. *Model specification*

The analysis of student drop-out behaviour has received much attention in the USA, where one of the most influential theoretical explanations of the attrition of students is the path analyses model of Tinto (1975, 1987). This model and related analyses suggest that the major determinants of completion are likely to be the student's

- (a) academic preparedness and
- (b) social and academic integration into the educational institution.

The analysis identifies several key influences on the probability of withdrawal, including the student's previous schooling, prior academic performance, family background and personal characteristics, as well as institutional characteristics. All the covariates in the model refer to the first year of the programme. In our model of student drop-out probabilities, we include control variables reflecting the student's prior academic preparedness, their social background

Table 3. Derived marginal effects ($\times 100$) on the conditional exit probability of withdrawal from the medical degree programme, logit models†

<i>Variable</i> ‡	<i>With university dummy variables (FE)—model 1</i>		<i>With university characteristics§§—model 2</i>		<i>Random university effects (RE)—model 3</i>	
	<i>Drop-out versus continuation (q₁) (1)</i>	<i>Drop-out versus completion (q₂) (2)</i>	<i>Drop-out versus continuation (q₁) (3)</i>	<i>Drop-out versus completion (q₂) (4)</i>	<i>Drop-out versus continuation (q₁) (5)</i>	<i>Drop-out versus completion (q₂) (6)</i>
Intercept	-2.248 (0.02)*	-4.512 (0.00)**	-1.055 (0.54)	-3.527 (0.02)*	-2.037 (0.05)*	-3.927 (0.00)**
Time trend††	-1.743 (0.00)**		-1.764 (0.00)**		-1.882 (0.00)**	
1986 year dummy	0.061 (0.73)	0.033 (0.82)	-0.662 (0.05)*	-0.212 (0.45)	0.055 (0.78)	0.033 (0.82)
Entry age (aged ≤ 18)						
Aged 19–21	-0.094 (0.65)	0.435 (0.05)*	-0.056 (0.79)	0.533 (0.04)*	-0.116 (0.61)	0.391 (0.03)*
Aged > 21	1.035 (0.11)	2.776 (0.02)*	1.153 (0.09)	3.428 (0.02)*	0.965 (0.06)*	1.251 (0.00)**
Sex—male	0.346 (0.05)*	0.498 (0.00)**	0.367 (0.05)*	0.549 (0.00)**	0.374 (0.06)*	0.490 (0.00)**
British national	-0.834 (0.13)	0.625 (0.00)**	-0.860 (0.13)	0.667 (0.00)**	-0.783 (0.09)	1.029 (0.00)**
Non-UK fee student	-0.331 (0.50)	0.549 (0.31)	-0.459 (0.33)	0.675 (0.30)	-0.489 (0.43)	0.468 (0.14)
Accommodation (living at home)						
Campus	-0.253 (0.48)	0.417 (0.04)*	-0.223 (0.52)	0.365 (0.13)	-0.237 (0.52)	0.457 (0.09)
Other	0.470 (0.29)	0.674 (0.18)	0.494 (0.27)	0.396 (0.38)	0.476 (0.25)	0.424 (0.15)
School attended (local education authority)						
Grammar school	-0.460 (0.13)	-0.152 (0.50)	-0.496 (0.10)	-0.161 (0.55)	-0.590 (0.11)	-0.095 (0.71)
Independent school	0.101 (0.67)	-0.057 (0.76)	0.094 (0.69)	-0.112 (0.60)	0.156 (0.53)	-0.069 (0.72)
College of further education	0.217 (0.57)	-0.407 (0.07)	0.216 (0.57)	-0.451 (0.10)	0.230 (0.54)	-0.530 (0.19)
Other	0.036 (0.95)	0.747 (0.23)	0.022 (0.97)	0.658 (0.31)	0.099 (0.87)	0.514 (0.12)
Main entry qualifications						
A-level scores	-0.077 (0.01)**	-0.053 (0.02)*	-0.077 (0.01)**	-0.066 (0.01)**	-0.082 (0.01)**	-0.049 (0.03)*
Higher scores	-0.034 (0.61)	-0.058 (0.27)	-0.029 (0.67)	-0.059 (0.33)	-0.039 (0.61)	-0.049 (0.33)
Other A- or AS-level scores	-0.015 (0.54)	0.013 (0.49)	-0.018 (0.47)	0.004 (0.86)	-0.013 (0.61)	0.011 (0.55)
Other Higher scores	-0.173 (0.05)*	-0.084 (0.29)	-0.181 (0.04)*	-0.157 (0.09)	-0.193 (0.05)*	-0.088 (0.25)
Achieved top score	-1.064 (0.00)**	0.117 (0.62)	-1.046 (0.00)**	0.087 (0.75)	-1.158 (0.00)**	0.104 (0.63)
Has mathematics A-level or Higher	0.180 (0.54)	0.099 (0.67)	0.190 (0.53)	0.131 (0.63)	0.190 (0.56)	0.119 (0.61)

(continued)

Table 3 (continued)

Variable‡	With university dummy variables (FE)—model 1		With university characteristics§§—model 2		Random university effects (RE)—model 3	
	Drop-out versus continuation (q ₁) (1)	Drop-out versus completion (q ₂) (2)	Drop-out versus continuation (q ₁) (3)	Drop-out versus completion (q ₂) (4)	Drop-out versus continuation (q ₁) (5)	Drop-out versus completion (q ₂) (6)
No favoured subject	2.686 (0.05)*	-0.575 (0.01)**	2.741 (0.05)*	-0.695 (0.00)**	1.992 (0.01)**	-0.869 (0.18)
1 favoured subject	-0.016 (0.98)	-0.422 (0.14)	0.026 (0.97)	-0.458 (0.19)	-0.010 (0.99)	-0.588 (0.34)
3 favoured subjects	-0.527 (0.05)*	-0.339 (0.12)	-0.547 (0.05)*	-0.451 (0.08)	-0.586 (0.05)*	-0.315 (0.14)
Other entry qualifications (A-levels or Highers)						
UK university qualifications	-1.460 (0.00)**	0.442 (0.37)	-1.481 (0.00)**	0.420 (0.43)	-2.289 (0.00)**	0.328 (0.31)
Other	-1.993 (0.00)**	0.132 (0.82)	-1.995 (0.00)**	0.137 (0.84)	-3.883 (0.00)**	0.091 (0.86)
Parental social class (professional)						
Intermediate	-0.094 (0.71)	-0.381 (0.03)*	-0.091 (0.72)	-0.456 (0.03)*	-0.113 (0.68)	-0.396 (0.04)*
Skilled non-manual	0.756 (0.09)	-0.158 (0.52)	0.719 (0.11)	-0.245 (0.36)	0.700 (0.07)	-0.172 (0.56)
Other	0.310 (0.32)	-0.159 (0.40)	0.322 (0.31)	-0.200 (0.36)	0.309 (0.32)	-0.163 (0.45)
Father or guardian is a doctor	-0.215 (0.48)	-0.169 (0.43)	-0.207 (0.50)	-0.195 (0.44)	-0.224 (0.52)	-0.189 (0.45)
Medical school characteristics						
Number of undergraduates			0.001 (0.37)	0.002 (0.03)*		
Number of taught post-graduates			0.008 (0.05)*	0.004 (0.24)		
Number of research post-graduates			-0.010 (0.31)	-0.024 (0.01)**		
Salaries per medical student			0.047 (0.01)**	-0.005 (0.74)		
Other expenses per medical student			0.097 (0.33)	-0.042 (0.63)		
Percentage of all staff—professors			-0.146 (0.05)*	0.128 (0.04)*		
Percentage of all staff—senior			-0.028 (0.36)	-0.093 (0.01)**		
Percentage of all staff—research			0.011 (0.43)	0.028 (0.03)*		
Treatment of university effects	Fixed	Fixed	Excluded	Excluded	Random	Random
Likelihood ratio test for the variance of the RE = 0					30.12 (0.000)	54.96 (0.000)

(continued)

Table 3 (continued)

<i>Variable</i> [‡]	<i>With university dummy variables (FE)—model 1</i>		<i>With university characteristics§ §§—model 2</i>		<i>Random university effects (RE)—model 3</i>	
	<i>Drop-out versus continuation (q₁) (1)</i>	<i>Drop-out versus completion (q₂) (2)</i>	<i>Drop-out versus continuation (q₁) (3)</i>	<i>Drop-out versus completion (q₂) (4)</i>	<i>Drop-out versus continuation (q₁) (5)</i>	<i>Drop-out versus completion (q₂) (6)</i>
Maximized log-likelihood value	-2935.49	-507.77	-2941.62	-530.56	-2958.42	-529.82
Number of observations	7789		7789		7789	

† *p*-values are given in parentheses.

‡ For classes of dummy variables the default class is indicated in parentheses.

§ Model 2 additionally includes individual dummy variables for the University of London, Scottish universities, a Welsh university and an Irish university.

§§ When RE were added to model 2, the RE variance was insignificantly different from 0.

* Significance at the 5% level.

** Significance at the 1% level.

†† The likelihood ratio statistic for testing this specification in the model with separate time dummy variables was 0.22 (0.99) for model 1 and 0.24 (0.99) for model 2.

and personal characteristics. In an analysis of the drop-out behaviour of all UK university students matriculating in 1989, Smith and Naylor (2001b) found evidence in support of a role for both academic preparedness and social integration. Previous schooling includes both prior qualifications of students and the type of school that they attended before university. Part of the motivation for the latter comes from the general issue of the influence of the quality and type of school on later outcomes (see Moffitt (1996)). As discussed in the previous subsection, in one of our models we include dummy variables for the medical school that was attended (model 1). Because of small cell sizes and the amalgamation of some London medical schools over our sample period, the University of London medical schools were treated as a single medical school. In a second model (model 2), we replace the medical school dummy variables with a set of variables measuring characteristics of the medical school that was attended by the student. The third model presented is the RE version of model 1 (model 3).

To obtain data on a satisfactory number of drop-outs in each year of the programme, the models have been estimated by combining the 1985 and 1986 entry cohorts, with a year dummy included to account for any aggregate time effect. The dependent variable takes the value 1 in a year if the individual drops out of the programme in that year and 0 otherwise. Because of the very small number of withdrawals after year 4, the final year conditional exit probability refers to year 4 or year 5. We thus have one conditional exit probability model for years 1–3 (see q_1 in equation (2)), which models the drop-out or continuation process, and another specification for the rest of the period (q_2 in equation (4)), which models the drop-out or completion process.

Before discussing the results, we consider the choice of the distributional assumptions that are used for our models. Three types of model were used—extreme value, logit and probit. Among the distributions the probit performs poorly, although there is essentially no difference between the extreme value model and the logit model. The estimated marginal effects were very similar across the extreme value and the logit models. The restriction that the effects of the covariates are the same in both the drop-out probabilities q_1 and q_2 in the logit specification is easily rejected

by the data. On the basis of this, we choose the logit model with unrestricted coefficients across the two drop-out probabilities as our preferred model for both the FE and the RE models.

5. Empirical results

The derived marginal effects on the rates of withdrawal and the corresponding p -values for the estimated logit models are reported in Table 3. Model 1 includes a set of medical school indicator variables among the covariates. The results from model 1 are presented in columns (1) and (2). This is the FE model where the unobservable medical school effects are treated as fixed numbers. To investigate further the *ceteris paribus* effects of medical schools, as discussed earlier, we also experimented by replacing the dummy variables with some variables that reflected the nature of the medical faculty in that particular university. These were the number of undergraduates, the number of post-graduates on degrees taught, the number of post-graduates on research degrees, the expenditure on salaries per medical student, the expenditure per medical student from research grants, the percentage of professors among the staff in the faculty, the percentage of senior staff, the percentage of research staff and separate binary indicators for London, Scottish, Welsh and Irish universities. Results from substituting the medical school dummy variables with medical school characteristics, labelled model 2, are presented in columns (3) and (4). Columns (5) and (6) report results from the RE model where the medical school unobservable effects are treated as random drawings from a normal distribution and labelled model 3. Columns (1), (3) and (5) refer to the conditional drop-out *versus* continuation probabilities q_1 , and columns (2), (4) and (6) to the conditional drop-out *versus* completion probabilities q_2 .

First, we note that the estimated marginal effects from these three models are broadly similar. The restriction that the medical school characteristics can be used in the model in place of the binary medical school indicators in the q_2 -probability is rejected very easily ($\chi^2(6) = 45.6$ (0.000)), but in the q_1 -probability it is not rejected at the 5% significance level ($\chi^2(6) = 12.26$ (0.057)). We also find that the maximized value of the log-likelihood is highest for both q_1 and q_2 from model 1 (the FE model). One advantage of treating the unobservable medical school effects as random is that we can include medical school characteristics as additional covariates. But having included the medical school characteristics in model 3 (or equivalently medical school REs in model 2), we could not reject the hypothesis of a zero variance on the REs at all conventional significance levels, i.e. the additional medical school characteristics seem to wipe out completely all the unobservable medical-school-specific variations. Hence we present results only for model 3 without additional medical school characteristics and for model 2 without additional medical school REs. We first discuss the results from model 1 and then highlight the differences between these results and those of model 2 and model 3.

Significant age effects are found only in q_2 . These suggest that conditionally on progressing to the final years of the programme, relative to someone who is 18 years old at the time of enrolment, an individual who is aged 22 years or more is estimated to have an increased withdrawal rate of 2.8 percentage points, *ceteris paribus*. Males are found to be significantly more likely than females to withdraw during all parts of the programme.

Fee status is not found to play a significant role in the drop-out behaviour, but nationality has a significant effect in the latter part of the programme. A student who is a British national is found to be 0.63 percentage points more likely to drop out during the latter part of the programme compared with someone who is a foreign national, *ceteris paribus*.

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 that were obtained for students who had taken either A-level or Higher

qualifications. 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 respectively three or five A-level or Higher passes before entry, some students do more subjects than the required number. To allow for this, we include the actual scores that they obtained in respectively their best three or five A-level or Higher examinations and also the scores from the rest, if they had more than the required number. As expected, the effect from A-levels on the probability of drop-out during the entire degree programme period is estimated to be negative and declining towards the final years of the programme. An extra 2 points on the A-level average (which is equivalent to an extra A-level grade) reduces the probability of drop-out by about 0.15 percentage points, *ceteris paribus*, during the earlier parts of the programme, and by about 0.11 percentage points in the final year of the programme. However, other A-level scores in addition to those which were achieved in the student's best three A-levels had no significant effect on the probability of drop-out. In contrast, the Highers score from the best five subjects is not found to have a significant effect, although students who have more than the five minimum required Highers are found to be significantly less likely to drop out in the early parts of the programme. In addition to the effect coming via the total A-level or Highers score, we also find that those students who respectively entered with the maximum possible score for their best three or five A-level or Highers subjects were significantly less likely to drop out during the early part of the programme. The finding that the student's level of performance at A-level or Highers has statistically significant effects on the probability of drop-out is in line with the hypothesis that academic preparedness for a medical degree is an important factor determining continuation and completion.

Looking at A-level or Higher subjects, assuming that a strong background in biology, chemistry and physics is likely to be appropriate for medical students, we include three dummy variables for those students who have none, one or three of these subjects. An original specification had dummy variables for whether the student had A-levels or Highers in physics, chemistry or biology or some combination of these. The results in Table 3 were arrived at having tested the implied restrictions. Relative to those with just two of the three of these subjects, students who enter not having done any of these subjects are found to be significantly more likely to drop out during the first few years of the programme, but this effect is completely reversed in the final year. As expected, students with all three of these subjects are found to have a lower probability of withdrawal relative to those with only two of the three subjects, *ceteris paribus*. Whether a student enters with only one or two of these subjects out of the three preferred subjects is not found to make any difference to the probabilities of withdrawal. Interestingly, there does not seem to be any additional effect coming via the possession of an A-level or Higher in mathematics.

We find that students who enter with qualifications other than the standard A-levels or Highers are significantly less likely to drop-out during the early parts of the programme. For example, students who had previously obtained a UK university qualification are more likely to continue beyond the third year. There was no significant effect in the final part of the programme.

It has been argued that the A-level performance of students from the non-independent-school sector might be an underestimate of their underlying ability and hence the negative effect of A-level performance might be lower for ex-pupils of state schools compared with independent schools (see, for example, Smith and Naylor (2001a)). To test for these kinds of effect, we included binary indicators for the type of school and also interactions of these with the actual scores that were obtained in the main entry qualification. We did not find any significant effect of A-level performance varying with the type of school, and therefore we report in Table 3 only those results that were obtained without these interactions. The only consistently significant

effect on probabilities of withdrawal associated with the type of school attended was for students who had attended a college of further education. This significant negative effect was only present in the final year of the programme.

The social class effects are found to be weak in general and show little effect. One exception is the significant negative effect for individuals from social class ‘intermediate’ (social class II) on the rate of withdrawal in the latter parts of the programme for students. We also note that a student whose parent is a medical practitioner is less likely to withdraw, although the effect is not well determined. We have also interacted prior academic performance with social class background, but we found no significant effects.

From model 1, the estimated coefficients on the medical school indicator variables permit the construction of a ranking of medical schools. In Fig. 1 we plot, for illustration, for the q_1 -probability, the estimated medical school effects (marked with a cross) along with their 95% confidence intervals. We find that only a few of the individual institution effects are significantly different from the median medical school, with five and one significantly above and below the median respectively.

In model 2 we replace the binary indicator variables for medical school attended with various characteristics of medical schools. The results for the model 2 specification are reported in columns (3) and (4) of Table 3. Although the estimated marginal effects in the q_1 -probability were broadly similar across the two models, there were some notable exceptions in q_2 . The absolute magnitude of the estimated marginal effects in the q_2 in model 2 was typically higher compared with model 1. Of the effects of the medical school characteristics themselves, we note from Table 3 that there is evidence that the probability of drop-out is lower the fewer students there are on the degree programme and the greater the number of post-graduate research students in the medical school.

We next turn to the results that are reported in columns (5) and (6) of Table 3. These refer to model 3 which is based on the RE specifications. First, we note that a test of hypothesis $H_0: \text{variance} = 0$ can be tested as a likelihood ratio test but the statistic will not be a standard χ^2 -statistic, since the parameter restriction is on the boundary of the parameter space. The

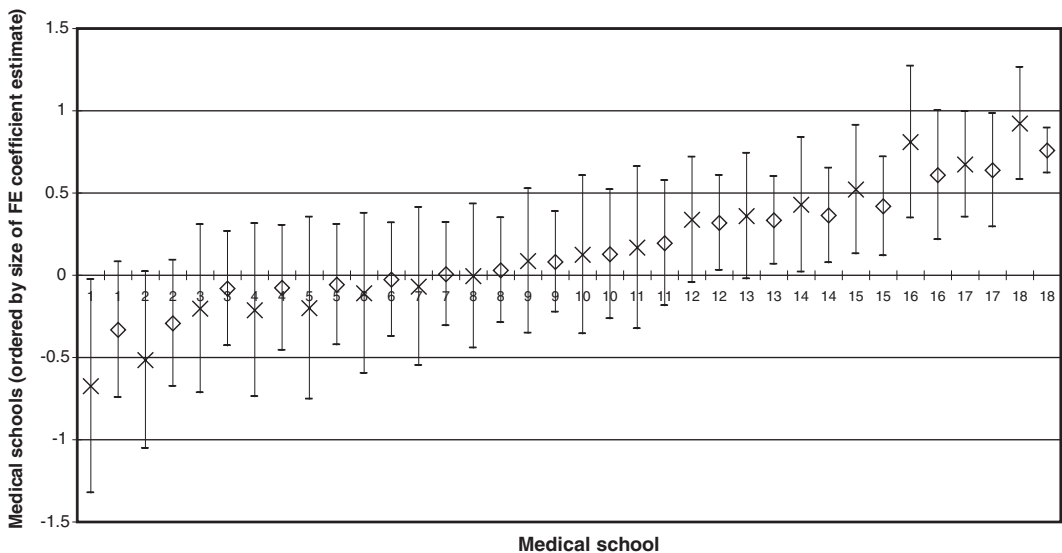


Fig. 1. Estimated medical school coefficients and 95% confidence intervals: x, FE; ◇, RE

standard likelihood ratio statistic has a probability mass of 0.5 at zero and $0.5\chi^2(1)$ for positive values. Thus a one-sided 5% significance level requires the use of the 10% critical value (Lawless, 1987). The likelihood ratio statistic for testing that the variance of the medical school REs is 0 is 30.12 in the model for q_1 and 54.96 in the model for q_2 , both of which are easily rejected at conventional levels of significance.

We turn next to the estimated effects of the covariates and simply note that the effects are very similar to the effects that were obtained from model 1. This should not be surprising since, because of our extremely large sample of the student population in each medical school, we would expect the RE and the FE estimates of the unobservable medical school effects to be the same.

As already noted, Fig. 1 plots the estimated medical school coefficients from the FE model (model 1) for q_1 . In addition, Fig. 1 also plots (for q_1 for the RE model (model 3)) the estimated shrunken residuals (empirical Bayes estimates) for each medical school (using the same order of medical schools as for the FE estimates) along with the 95% confidence intervals—marked with a diamond symbol. From the RE model, seven medical schools are significantly above the median, although none are significantly below. It is clear from Fig. 1 how similar the estimated medical school effects are from the two models (with a rank correlation in excess of 0.98), although the effects from the RE model tend to be estimated more precisely.

6. A competing risks analysis

The models that have been considered so far are for the conditional probability of withdrawal from the programme, conditioned on not having withdrawn up to that point. These are ‘single-risk’ models, in the sense that no distinction is made between various alternative reasons for withdrawal. A student can withdraw from the programme for reasons such as transfers between courses, academic failure and health problems, *inter alia*. The fourth and fifth columns of Table 1 provide the breakdown of the number of drop-outs by reason for withdrawal. Since in our model specification the continuation *versus* drop-out probabilities q_1 was specified over the first 3 years of the programme, we report the relevant frequencies for this part of the programme only. We note that there were no withdrawals due to changes in course in years 4 or 5. Of those students who dropped out of the medical programme, the proportion of students who changed their course of study was about a third. From the point of view of public policy makers seeking to expand the number of fully trained medical doctors, distinguishing between reasons for dropping out may not be important: a drop-out is a drop-out. Against this, strategies for minimizing the rate of drop-out might benefit from a better understanding of the nature of the decision to drop-out. Additionally, for the university in which the medical school is located, the distinction between transferring course and dropping out of the medical programme for other reasons may be particularly important if only the latter impacts adversely on the institution’s performance indicator score.

In this section, since the FE model (model 1) produced the highest maximized log-likelihood value, we extend this model to the competing risks case where we distinguish between the reported reasons for the exit. See Narendranathan and Stewart (1993b) for an example of a hazard model with competing risks in the context of exiting from unemployment to employment. In the current paper, given the small number of exits, we consider only two possible alternative reasons for exit: a ‘course transfer’ or ‘other reasons’. We wish to model the determinants of the conditional probability of exiting out of the medical programme for one of these two reasons, assuming that the exits occur at the start of the interval.

Two separate logit models were fitted where, in the model for drop-out due to other reasons, a course transfer is treated as a right-censored event, and vice versa. The coefficient estimates for

the conditional probability of exiting out of the medical programme for other reasons (model 4a) and the conditional probability of exiting out of the medical programme because of course transfer (model 4b) are presented in columns (2) and (3) of Table 4 respectively. For ease of comparison, the coefficient estimates from the single-risk model (model 1) are reported in column (1).

With respect to personal characteristics, age and gender effects are found to be significant only in the conditional probability for exiting due to reasons other than course transfers. In contrast, the nationality and fee status variables are found to exert a significant influence only on the course transfer conditional probability relative to progression. Unlike in the single-risk model (model 1), a student who attended a grammar school compared with someone coming from a local education authority school, before entry into medical school, is found to be significantly less likely to exit for other reasons, *ceteris paribus*.

With respect to prior qualifications, the total A-level scores in the best three subjects is found to play an important and significant role only in the exit probability that is associated with other reasons, although having achieved a top score has a significant effect on both. We note that students with no favoured A-level subjects are significantly more likely to drop out for reasons other than course transfer. In contrast, having three, rather than two, favoured subjects significantly reduces the probability of transferring course: we infer from this that matching seems to be important.

The significant positive effect that was found for students coming from a 'skilled non-manual' social class background in model 1 is now found to act via the risk that is associated with withdrawals for other reasons. Interestingly, we also find that students coming from a medical background are no more or less likely to withdraw from the medical school but are significantly less likely to transfer course. This result is consistent with the idea that such students have a greater commitment to their chosen course of medical study.

7. Discussion and further remarks

In the context of a shortage of trained doctors in the UK, the Government's policy for the medical workforce rests on a plan to implement an on-going major expansion in total medical school intake (see Medical Workforce Standing Advisory Committee (1997) and Higher Education Funding Council for England (2001)). As part of the strategy of expansion, the Medical Workforce Standing Committee has called for research into the factors affecting the progression and withdrawal behaviour of medical students. A substantial body of previous work on withdrawals of UK university students has exploited data on full cohorts of students but has omitted any analysis of medical students as it is generally accepted that, for a variety of reasons, the characteristics of medical students and their courses differ from those of other students and hence justify a separate analysis. In this paper, we conducted a statistical analysis of the drop-out behaviour of UK medical students, exploiting data on full student cohorts.

Our statistical analysis extends previous research in the general area by adopting a hazard model to model the probability of dropping out. We also offer a competing risks framework. The exit probability model is based on the initial distinction between dropping out and continuing and the subsequent distinction between dropping out and completing, conditional on continuation. For each of our single-risk models, we first considered two alternative specifications: in one, including binary indicator variables for the particular medical school attended (the FE model) and, in the other, incorporating medical school characteristics. We also re-estimated our single-risk model, treating the unobserved medical school effects as random, and found the results to be broadly similar to those of the FE model. The competing risks analysis

Table 4. Maximum likelihood estimates: single- and competing risks logit models†

<i>Variable‡</i>	<i>Model 1—single risk, drop-out versus continuation (q₁) (1)</i>	<i>Model 4a—competing risks, drop-out for reasons other than course transfers versus continuation (2)</i>	<i>Model 4b—competing risks, drop-out for reasons of course transfers versus continuation (3)</i>
Intercept	-0.973 (0.02)§	-1.676 (0.00)§§	-2.050 (0.00)§§
Time trend	-0.754 (0.00)§§	-0.638 (0.00)§§	-1.021 (0.00)§§
1986 year dummy	0.026 (0.74)	-0.085 (0.36)	0.291 (0.04)§
Entry age (aged ≤ 18)			
Aged 19–21	-0.041 (0.65)	0.022 (0.84)	-0.224 (0.18)
Aged > 21	0.380 (0.06)	0.669 (0.00)§§	-0.224 (0.18)
Sex—male	0.151 (0.06)	0.181 (0.06)	0.075 (0.60)
British national	-0.315 (0.09)	-0.083 (0.71)	-0.924 (0.00)§§
Non-UK fee student	-0.153 (0.53)	0.153 (0.57)	-1.099 (0.05)§
Accommodation (living at home)			
Campus	-0.106 (0.46)	-0.115 (0.50)	-0.035 (0.90)
Other	0.190 (0.25)	0.281 (0.13)	-0.249 (0.51)
School attended (local education authority)			
Grammar school	-0.214 (0.16)	-0.342 (0.07)	0.061 (0.81)
Independent school	0.043 (0.66)	0.000 (1.00)	0.140 (0.43)
College of further education	0.091 (0.55)	0.080 (0.66)	0.117 (0.65)
Other	0.016 (0.95)	-0.085 (0.76)	0.336 (0.46)
Main entry qualifications			
A-level scores	-0.033 (0.01)§§	-0.035 (0.02)§	-0.020 (0.42)
Higher scores	-0.015 (0.61)	-0.038 (0.27)	0.031 (0.56)
Other A- or AS-level scores	-0.006 (0.54)	-0.019 (0.14)	0.016 (0.35)
Other Higher scores	-0.075 (0.05)§	-0.050 (0.21)	-0.099 (0.10)
Achieved top score	-0.506 (0.00)§§	-0.499 (0.00)§§	-0.530 (0.01)§§
Has mathematics	0.079 (0.55)	0.150 (0.34)	-0.052 (0.82)
A-level or Higher			
No favoured subject	0.804 (0.01)§§	0.974 (0.00)§§	0.250 (0.75)
1 favoured subject	-0.007 (0.98)	-0.015 (0.97)	0.083 (0.89)
3 favoured subjects	-0.227 (0.06)	-0.146 (0.32)	-0.363 (0.08)
Other entry qualifications (A-levels or Highers)			
UK university qualifications	-0.920 (0.00)§§	-0.836 (0.00)§§	-2.825 (0.00)§§
Other	-1.588 (0.00)§§	-1.538 (0.00)§§	-2.825 (0.00)§§
Parental social class (professional)			
Intermediate	-0.041 (0.71)	0.077 (0.57)	-0.250 (0.17)
Skilled non-manual	0.291 (0.06)	0.470 (0.01)§§	-0.094 (0.73)
Other	0.129 (0.30)	0.224 (0.15)	-0.086 (0.68)
Father or guardian is a doctor	-0.096 (0.50)	0.160 (0.33)	-0.700 (0.01)§§
University dummy variables included	Yes	Yes	Yes
Maximized log- likelihood value	-2935.48	-2204.29	-1077.32
Proportion who exit into given state during the early part of the programme	0.091	0.063	0.028

†*p*-values are given in parentheses. Because of a lack of enough observations to identify the effects, the model presented in column (3) has two variables entered as a sum. These were on 'other entry qualifications' and the age dummy variable.

‡For classes of dummy variables the default class is indicated in parentheses.

§Significance at the 5% level.

§§Significance at the 1% level.

distinguishes between dropping out because of course transfer and dropping out for all other reasons.

Our main finding is that academic preparedness is the major influence on the withdrawal of medical students. Students who scored highly in their pre-university qualifications (at A-level or Highers, for example) are significantly less likely to drop out of medical school: they are both more likely to continue and more likely to complete than students performing less well before entering university. We also find that the choice of pre-university subjects exerts a significant influence over the medical student's drop-out probability. For example, students who had taken chemistry, physics and biology at A-level were significantly less likely to drop out than were students with only two (or fewer) of these three subjects. In particular, students with all three of chemistry, physics and biology were significantly less likely to transfer from medicine to some other course: we infer that the extent of the pre-university subject match is an important characteristic of medical students. This is not true for all university degree subjects (see Smith and Naylor (2001b)). On other entry qualifications, we find that medical students who had already obtained a UK university qualification are significantly less likely to drop out of medical school: this could have implications for a lower average drop-out rate on recently introduced post-graduate medical degree programmes.

Among other results, we also find that age and gender exert significant influences on drop-out behaviour, with males and more mature students more likely to withdraw than females or younger students respectively. Conversely, we find surprisingly little effect associated with either school or family social class background, although we do note that students with a medical doctor as a parent are significantly less likely to transfer out of a medical programme. We find that there are some significant differences in drop-out probabilities across the medical schools, after controlling for the students' observed characteristics. In the specifications based on medical school characteristics, we find evidence that the drop-out probability is lower in smaller medical schools and in medical schools with a high proportion of post-graduates on a medical programme. In terms of the debate on widening access into medical school for students who do not meet the traditional levels of attainment in their pre-university qualifications, our results raise the serious concern that, potentially, the *ex ante* completion probability of such students will be relatively low. This is not to argue against policies of widening access. It is to point out, however, that such policies may be most successfully accomplished if they are accompanied by complementary strategies of student support (including appropriate supplementary training) and by well-targeted recruitment activity. With respect to the latter, it has been shown that, for non-medical students, prior levels of attainment might underpredict university performance for students from certain backgrounds. For example, students who previously attended state schools tend to perform better at university than students from the independent school sector, *ceteris paribus* (see Smith and Naylor (2001a)). Accordingly, in our current study of medical student drop-out behaviour we have included controls both for type of previous school and for potential interactions between prior qualification attainment and type of school. We found no clear evidence of significant effects. Nor do we find significant interaction effects associated with family background and prior qualifications. It follows that, if targeted recruitment to widen access is to be successful in minimizing the effect on the drop-out probability, targeting may need to be more precise than simply making differential offers according to the type of school, as has begun to occur elsewhere in the UK higher education sector.

Acknowledgements

We are grateful to Norman Ireland, Chris McManus, Costas Meghir, Geert Ridder, Jim Walker,

seminar participants at the Institute for the Study of Labour, Bonn, at the University of Warwick and the Centre for Health Service Studies. Comments from two referees and the Joint Editor are also gratefully acknowledged. We acknowledge both the Universities' Statistical Record, as the original depositors, and the UK Data Archive for the use of data set SN:3456 of the Universities' Statistical Record. None of these individuals or organizations bears any responsibility for any of the analysis or interpretations that are presented in this paper.

References

- Angel, C. and Johnson, A. (2000) Broadening access to undergraduate medical education. *Br. Med. J.*, **321**, 1136–1138.
- Arulampalam, W., Naylor, R. A. and Smith, J. (2003) Factors affecting the probability of first-year medical student dropout in the UK: a logistic analysis for the intake cohorts of 1980–1992. *Med. Educ.*, to be published.
- 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. *J. R. Statist. Soc. A*, **158**, 297–318.
- Goldstein, H. (2003) *Multilevel Statistical Models*, 3rd edn. London: Arnold.
- Higher Education Funding Council for England (2001) Increasing medical student numbers in England. *Report*. Higher Education Funding Council for England, Bristol.
- Lancaster, T. (1990) *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press.
- Lawless, J. F. (1987) Negative binomial and mixed Poisson regression. *Can. J. Statist.*, **15**, 209–225.
- Machin, S. and Oswald, A. (2000) UK economics and the future supply of academic economists. *Econ. J.*, **110**, F334–F349.
- McManus, I. C. (1996) Drop-out rate in medical school seems reasonable. *Br. Med. J.*, **313**, 173.
- McManus, I. C. (1998) Factors affecting likelihood of applicants being offered a place in medical schools in the United Kingdom in 1996 and 1997: retrospective study. *Br. Med. J.*, **317**, 1111–1117.
- McManus, I. C., Richards, P. and Winder, B. C. (1999) Intercalated degrees, learning styles, and career preferences: prospective longitudinal study of UK medical students. *Br. Med. J.*, **319**, 542–546.
- Mealli, F., Pudney, S. and Thomas, J. (1996) Training duration and post-training outcomes: a duration-limited competing risks model. *Econ. J.*, **106**, 422–433.
- Medical Workforce Standing Advisory Committee (1997) Planning the medical workforce: third report. *Report*. Department of Health, London. (Available from <http://www.doh.gov.uk/medical/mwsca3.htm>.)
- Moffitt, R. (1996) Symposium on school quality and educational outcomes: introduction. *Rev. Econ. Statist.*, **78**, 559–561.
- Narendranathan, W. and Stewart, M. B. (1993a) How does the benefit effect vary as unemployment spells lengthen? *J. Appl. Econometr.*, **8**, 361–381.
- Narendranathan, W. and Stewart, M. B. (1993b) Modelling the probability of leaving unemployment: competing risks models with flexible base-line hazards. *Appl. Statist.*, **42**, 63–83.
- Parkhouse, J. (1996) Intake, output, and drop out in United Kingdom medical schools. *Br. Med. J.*, **312**, 885.
- Porter, O. (1990) *Undergraduate Completion and Persistence in Four-year Colleges and Universities*. Washington DC: National Institute of Independent Colleges and Universities.
- Rabe-Hesketh, S., Pickles, A. and Skrondal, A. (2001) GLLAMM manual. *Technical Report 2001/01*. Department of Bio-statistics and Computing, Institute of Psychiatry, King's College, London.
- Smith, J. P. and Naylor, R. A. (2001a) Determinants of individual degree performance. *Oxf. Bull. Econ. Statist.*, **63**, 29–60.
- Smith, J. P. and Naylor, R. A. (2001b) Dropping out of university: a statistical analysis of the probability of withdrawal for UK university students. *J. R. Statist. Soc. A*, **164**, 389–405.
- Stata Corporation (2001) *Statistical Software*. College Station: Stata Corporation.
- Tinto, V. (1975) Dropout from higher education: a theoretical synthesis of recent research. *Rev. Educ. Res.*, **45**, 89–125.
- Tinto, V. (1987) *Leaving College: Rethinking the Causes and Cures of Student Attrition*. Chicago: University of Chicago Press.
- Van Ours, J. C. and Ridder, G. (2000) Fast track or failure: a study of the completion rates of graduate students in economics. *Discussion Paper 2363*. Centre for Economic Policy Research, London.