

Identifying Medical School Applicants from Ethnic Minorities

I. C. McMANUS, S. L. MAITLIS & P. RICHARDS

St Mary's Hospital Medical School, Imperial College of Science, Technology and Medicine, London

ABSTRACT *Two studies of applicants to United Kingdom medical schools show that ethnic origin of surnames is reliably assessable by independent judges, and that surnames are valid indicators of ethnic origin as determined by self-classification, showing very high specificity (97%) and slightly lesser sensitivity (84%). Ethnic origin can also be determined from residential information derived from post-codes and place of birth, information in each case being highly specific (99% and 98%) but lacking in sensitivity (25% and 33%). The addition of place of birth and post-code data to surnames provides an increase in overall sensitivity (90%) with no improvement in specificity (94%). A comparison of survey respondents and non-respondents shows that applicants from ethnic minorities are somewhat less likely to respond than non-minority applicants, although the effect is small. Responding applicants from ethnic minorities reply as quickly as non-minority applicants. Our survey confirms the feasibility of direct monitoring of the ethnic origin of applicants by asking applicants to complete a short questionnaire, and of its indirect monitoring by the use of surname. Both UCCA (Universities' Central Council on Admissions) and PCAS (Polytechnic Central Admissions System) have announced that they will instigate the ethnic monitoring of applicants for admission in autumn 1990.*

Introduction

The Race Relations Act of 1976 makes it illegal to discriminate against individuals on the basis of membership of a 'racial group', defined within Section 3 of the Act in terms of distinctiveness of colour, race, nationality or ethnic or national origin (Parekh, 1987). If discrimination is to be adequately proven or refuted it is necessary to have a satisfactory method of identifying applicants from ethnic minorities who apply to medical school. That, however, has not been possible using any of the information which is currently available on the UCCA application form, with the exceptions of place of birth and nationality (the latter of which is probably a special case in view of the separate regulations regarding fees for such students). The Commission for Racial Equality (CRE) recommends that equal opportunity policies should be monitored, and that the first stage of such monitoring is the classification of individuals by ethnic origin (CRE, 1984). The need for such routine identification of minority applicants has increased since it emerged that one London medical school was inadvertently discriminating against applicants from ethnic minorities (Smith, 1987; CRE, 1988; Lowry & Macpherson, 1988).

Collier & Burke (1986; see also Veitch, 1984) made a claim, without providing evidence for its reliability or validity, that applicants from ethnic minorities can be identified on the

basis of having non-European surnames, although they noted that applicants of West Indian origin were often missed by such a technique. McManus & Richards (1985) classified the surnames of all applicants in the 1981 cohort of applicants to St Mary's Hospital Medical School and noted a reasonable correlation between non-European surnames and membership of an ethnic minority determined from photographs of those applicants interviewed.

Applicants to university are not currently asked to describe their ethnic origin on the UCCA form, and both UCCA and the Committee of Vice-chancellors and Principals (CVCP) have, until recently, shown considerable reluctance to ask applicants for such information, even for statistical purposes. In addition there is some indication that applicants may be wary of giving ethnic self-descriptions because of fears that the information may itself be used in a discriminatory fashion. It has not been possible, therefore, to validate alternative measures for identifying minority applicants using internal evidence. Further, neither UCCA nor any other body has been in a position to assess patterns of application for putative discrimination using the information available on the UCCA form.

Only three pieces of data on the UCCA form might be used to identify successfully an applicant's ethnic origin. The surname is the most likely. A second potentially useful source is the post-code which is routinely provided by applicants. Since minority applicants are more likely to live in areas of the country where other minorities live, small area census statistics derived from the post-codes (a straightforward process) could in principle identify minority applicants, at least on a statistical basis, and, in particular, might be useful in identifying minority applicants with typically European surnames. Information derived from post-codes (Speller & Hale, 1985) has already proved useful in other areas of health-related research (Morgan, 1983; Morgan & Chinn, 1983). A third possible source of information concerns the applicant's place of birth, although this measure might be expected to be of little use if most applicants from ethnic minorities are born in Britain (i.e. they are second- or third-generation Britons).

The present paper seeks to answer several types of question about identifying the ethnic origins of applicants.

(i) Can surnames be reliably categorised by independent raters both as European/non-European and also into more detailed subgroups?

(ii) Is there evidence that the rate of application of ethnic minority applicants is different at different medical schools, and, in particular, to what extent are applicants to St Mary's, the one school which has carried out comprehensive surveys of patterns of application, typical of applicants in general?

(iii) How valid are surnames as measures of ethnic origins when compared with descriptions of ethnic origin made by applicants themselves?

(iv) How useful are small area census statistics derived from applicants' post-codes in identifying ethnic origins, and can they provide additional information over and above that provided by surnames?

(v) Is place of birth effective in identifying the ethnic origin of applicants?

(vi) Are minority applicants less likely to respond to surveys of applicants, and if so, what is the extent of bias induced in surveys?

Two separate studies have been carried out, one covering all applications to UCCA to study medicine for admission in October 1984, and the other covering all applicants to St Mary's Hospital Medical School for admission in October 1986.

Study 1

Method

UCCA provided a computer-readable magnetic tape containing a brief summary of all applicants for admission to university in October 1984. Of these, 10,607 had included an application for medicine as at least one of their five choices, and these applications were used as the basis of the present study. Surnames and code numbers were transferred to a small microcomputer which was programmed to present a single surname at a time on the computer's visual display. The person classifying the names then pressed one of six keys to indicate whether the origin of the name was regarded by them as British, Other European, Middle Eastern or Arabic, from the Indian subcontinent, from the Far East, or from Africa. In addition a category was also available for Don't know/Other, although its use was heavily discouraged, and in fact it was not actually used during the study. Two individuals (ICM and SLM) separately rated each of the names according to its origin. ICM is a medically qualified Senior Lecturer in Psychology applied to Medicine who has researched medical student selection for a number of years, and who has also been involved in interviewing and selecting medical students. SLM is a recent graduate in psychology who had been working for the previous six months on a research project studying the 1986 cohort of applicants to St Mary's. Both have travelled in the Far East, and ICM has also travelled in the Middle East, the Indian subcontinent and Africa.

Results

Reliability of raters. Table I cross-classifies the judgements made by the two raters. Using the six separate categories of surname origin, there was agreement between the two raters in 9379 cases (88.4%). This compares with the 51.7% agreement which would be expected on the basis of a consideration of the distribution of marginal proportions. Comparison of the marginal proportions suggests that ICM used the Indian category nearly twice as frequently as did SLM, and used the Arab category less than did SLM. The major discrepancies between the two raters within the body of the table concerned the 297 individuals classified as Indian by ICM and Arab by SLM, and the 178 individuals classified as Indian by ICM and African by SLM.

TABLE I. Classification of ethnic origins of surnames of 10,607 applicants for admission to British medical schools in October 1984 by two individuals, ICM and SLM. The vertical and horizontal lines indicate the broad separation into European and non-European names. Diagonal elements in italics indicate complete agreement between the two raters.

SLM	ICM						Percentage
	British	European	Arab	Indian	Far Eastern	African	
British	7288	59	14	65	11	9	70.2
European	124	379	6	93	19	13	6.0
Arab	40	35	319	97	14	19	6.8
Indian	19	7	29	722	19	7	7.6
Far Eastern	35	3	2	30	515	10	5.6
African	16	16	22	178	17	156	3.8
Percentage	70.9	4.7	3.7	13.1	5.6	2.0	

Analysis of the two broad categories of European and non-European showed agreement between ICM and SLM in 10,206 cases (96.2%), with almost equal marginal proportions being coded as non-European by ICM (24.4%) and SLM (23.8%). The chance expectation of agreement for these marginal proportions is 63.4%. Using the criterion of a non-European surname as being one coded as non-European by *either* rater, then 26.0% of applicants had non-European surnames, compared with 22.2% on a criterion where *both* assessors rated the name as non-European. Since these criteria are almost equivalent, we shall, for the remainder of the study, confine ourselves to the more inclusive criterion, that a surname is non-European if either rater has coded it as non-European.

Variation between medical schools. Table II shows the proportion of applicants to each of the medical schools who did not hold UK nationality. Considering only the UK nationals, the table also shows the proportions who had non-European surnames, as well as the proportions who were female or who were mature applicants (defined as aged over 21 on 30 September 1984), these latter measures being derived also from the UCCA file. Medical schools are identified only by a random code. The median proportion of applicants with non-European surnames was highest in London (20.5%), next highest in medical schools of England and Wales (17.6%) and least in schools of Scotland and Northern Ireland (14.0%).

Of the 30 British medical schools, St Mary's was in the top third for numbers of applicants. Its proportions of non-UK applicants, and of UK applicants with non-European surnames, were both in the middle third of the overall range. St Mary's was in the bottom third for female applicants and the top third for mature applicants.

Demographic factors related to surname. The proportions of applicants by sex differed within different surname groups. Using ICM's classification of surnames among UK nationals, the proportion of females was highest in those with British (46.7%) and European (42.2%) surnames, intermediate in applicants with Middle Eastern/Arabic (38.5%), African (37.1%) and Indian (33.6%) surnames, and lowest of all in those with Far Eastern surnames (25.9%) ($\chi^2=101.6$, 5 df, $p<0.001$). Similar results were found with SLM's classification ($\chi^2=105.8$, 1 df, $p<0.001$). Surname groups also differed in the proportion of mature applicants ($\chi^2=40.3$, 5 df, $p<0.001$); using ICM's classification of UK nationals, the highest proportion of mature applicants was in those with Far Eastern surnames (23.9%), with similar proportions amongst the other ethnic groups: Indian (14.7%), European (13.9%), Arab (12.7%), British (12.0%) and African (11.3%).

Discussion

Two raters were highly reliable in their judgement of whether an applicant had a European or a non-European surname, agreeing in 96.2% of cases. They were less reliable in distinguishing specific origins for non-European surnames, with particular confusion arising between Indian and Arabic names, and between Indian and African names. Nevertheless there was still agreement in 72.6% of such cases, suggesting that useful information concerning specific origin can be extracted from surname analysis.

The proportions of UK applicants with non-European surnames differed between medical schools, with a general trend for London schools to have more applicants with non-European surnames than schools elsewhere in England and Wales, which in turn had more applicants than schools in Scotland and Northern Ireland. These patterns may well reflect

TABLE II. Proportions of applicants to each British medical school who are not UK nationals, and amongst those who are UK nationals, the proportions with non-European surnames (NES), who are female, and who are mature applicants

School	All applicants		UK applicants		
	N	Non-UK (%)	NES (%)	Female (%)	Mature (%)
London					
1	1821	13.8	21.0	46.7	16.3
2	2862	13.8	20.8	44.3	8.2
3	2115	16.0	17.6	42.3	15.1
4	1677	12.9	20.8	42.9	14.2
5	2985	15.5	19.6	42.7	13.4
6	3165	19.1	20.5	43.3	14.3
7	1789	15.2	19.6	42.6	12.7
8	855	12.2	15.0	39.0	11.3
9	941	17.9	20.3	38.4	12.4
10	1930	17.0	20.9	42.1	11.8
11	764	23.8	21.1	34.0	15.3
England and Wales					
12	2011	12.0	15.5	47.0	14.0
13	1894	14.5	18.3	44.6	9.0
14	717	12.4	14.6	44.0	9.1
15	1680	14.2	17.8	45.5	5.9
16	1848	14.0	17.6	46.6	12.2
17	1743	15.4	19.4	49.1	11.1
18	1510	18.2	16.9	46.8	29.8
19	1417	15.3	20.8	45.8	16.9
20	1085	16.0	13.6	48.1	22.5
21	1842	12.9	17.6	46.3	12.0
22	1740	11.8	14.3	45.4	6.0
Scotland and Northern Ireland					
23	1221	23.7	16.3	46.2	12.7
24	1064	15.9	13.2	45.6	10.6
25	718	13.8	14.9	44.9	8.6
26	1268	14.2	14.9	43.9	13.1
27	1300	19.4	11.2	44.8	8.6
28	575	17.7	7.8	37.4	10.1
Oxford and Cambridge					
30	481	12.5	13.8	43.5	1.7
31	832	13.0	15.7	40.9	2.3

differential patterns of settlement by recent immigrants to Britain (CRE, 1985), coupled with a tendency for applicants to apply to universities near to their home town.

St Mary's Hospital Medical School was fairly typical of medical schools in general; although it had one of the largest numbers of applications, and had, in terms of proportion, rather fewer women and rather more mature students, it had typical proportions of non-UK applicants and applicants with non-European surnames. We conclude that applicants to St Mary's are sufficiently representative of applicants in general for our cohort studies to be valid for extrapolation to British medical schools in general.

Study 2

Study 1 has shown that non-European surnames can be identified *reliably* (i.e. different raters agree on the classification of surnames), but it cannot show that non-European surnames are a *valid* measure of ethnic origin, since there is no criterion measure of ethnic origin. In study 2 we surveyed all candidates who applied to St Mary's Hospital Medical School for admission in October 1986, and obtained a criterion measure of ethnic origin in terms of the applicants' self-descriptions of ethnic origins.

Method

All those applying to St Mary's Hospital Medical School during the autumn of 1985 for admission in October 1986 were included in the 1986 cohort survey if they applied before the closing date of 15 December 1985 and had an address for correspondence within the British Isles. Within a few days of receipt of the UCCA form each applicant was sent a detailed questionnaire, 16 A4 pages in length, covering a range of topics concerning social, family, educational and demographic origins, interests, attitudes and moral reasoning. Within this questionnaire, on page 3, was half a side devoted to ethnic origins, the full text of which is shown in the Appendix. The questionnaire had an introduction to make clear the reasons for obtaining such information, before asking about the physical appearance of the applicant and the applicant's parents, about the place of birth of the applicant's parents and grandparents, the age at which the applicant started to learn English, and about the applicant's response to the questionnaire itself. In the event of applicants not returning the questionnaire two further follow-ups were sent, one after about four weeks and the other after eight weeks. It was emphasised to all candidates that the questionnaire was not a formal part of their application, and that they were under no obligation to complete it, but that the validity of the questionnaire research depended upon the proportions of applicants who returned questionnaires.

Surnames of applicants were classified by two individuals, ICM and DNJL, using the computer program described in study 1. ICM also took part in study 1; DNJL only took part in study 2: she is a medical registrar with an interest in medical education, who has travelled extensively in the Middle and Far East, in the Indian subcontinent and in Africa.

Post-codes for applicants were obtained from their UCCA forms. Small area census statistics and ACORN codes (CACI, 1985) were provided for each post-code by CACI, an agency specialising in the linking of census and other publicly available databases. For each post-code the program converted the code into an Ordnance Survey grid reference which was then related to the enumeration districts used for the 1981 census. For each such 'small area', information is available on 40 different measures concerning the immediate locale in which the post-code is situated. It should be emphasised that the post-codes do *not* provide information on a specific individual, but only on the aggregate of the 500 or so individuals who live within the enumeration district. Thus, for instance, if these data show a high rate of unemployment in an area it does not mean that the individual applicant or a member of their family is unemployed, merely that a high proportion of the people who live geographically close to the applicant are unemployed. The description is therefore of the *neighbourhood* and not of the individual; in so far as the characteristics of neighbourhoods are relatively homogeneous, it is possible to infer the likely social composition of individuals on these measures.

Place of birth information was available routinely for all applicants from the UCCA forms, and was analysed in two ways: by dividing applicants into those born in the UK and those born elsewhere, and by further distinguishing those applicants born in the New

Commonwealth (i.e. ex-colonies of Britain, but excluding Australia, New Zealand, Canada and other places where the majority of the population is ethnically white).

Results

Those who applied for admission to St Mary's amounted to 2399 individuals, and of these 2210 satisfied the criteria for inclusion in the 1986 St Mary's cohort. Those returning the postal questionnaire numbered 2050 (92.7%), and of these 2030 (99.0%) answered the questions about ethnic origins. 1456 (71.0%) described themselves as 'White, European', 362 (17.7%) as 'Asian', 77 (3.8%) as 'Far Eastern', 49 (2.4%) as 'Arab/Middle Eastern', 36 (1.8%) as 'African' and 15 (0.7%) as 'Caribbean'. In addition, 4 (0.2%) described themselves as 'Other' (2 such 'South American', 1 as 'Pathan' and 1 giving no other information) and 31 (1.5%) said they were of mixed origins, in all but one case having one parent described as 'White, European'. For convenience, applicants will be divided into categories of White and non-White, the latter referring to all self-classified ethnic groups except 'White, European'.

Both parents had been born in the UK in the case of 1174 applicants (57.3%), and all four grandparents had been born in the UK in the case of 1047 applicants (51.1%). Applicants who described English as being their mother tongue numbered 1800 (88.7%). Only 46 applicants (2.3%) reported difficulties in answering the ethnic questions, and these difficulties mostly reflected either confusion between physical appearance and nationality, or individual ignorance of ancestry. Only 4 applicants (0.2%) stated explicit disapproval of the substance of the questions. Applicants in favour of monitoring ethnic origins numbered 1587 (78.2%), almost equally divided between 'Probably' (41.0%) and 'Definitely' (41.1%). 195 applicants (10.1%) were 'Probably not' in favour, and 150 (7.8%) were 'Definitely not' in favour of such monitoring.

Reliability of assessment of surnames. Table III shows the agreement between the two raters in assessing the ethnic origins of surnames. There was complete agreement in 2149 cases (89.6%) and agreement over European/non-European in 2305 cases (96.1%), both values being similar to those found in study 1.

TABLE III. Classification of ethnic origins of surnames of 2399 applicants for admission to St Mary's Hospital Medical School in October 1986 by two individuals, ICM and DNJL. The vertical and horizontal lines indicate the separation into European and non-European names. Diagonal elements in italics indicate complete agreement between the two raters.

DNJL	ICM						Percentage
	British	European	Arab	Indian	Far Eastern	African	
British	<i>1547</i>	7	1	10	7	0	65.5
European	28	<i>100</i>	4	23	4	2	6.7
Arab	2	1	<i>76</i>	27	1	0	4.5
Indian	22	5	37	<i>271</i>	6	8	14.5
Far Eastern	0	1	1	2	<i>115</i>	2	5.0
African	7	5	8	21	8	<i>40</i>	3.7
Percentage	66.9	5.0	5.3	14.8	5.9	2.2	

Validity of assessment of ethnic origin using surnames. Table IV shows the agreement between each rater's assessment of the ethnic origin of a surname and the applicant's own classification of ethnic origin. There is complete agreement for 1721 (84.7%) of ICM's judgements and 1732 (85.3%) of DNJL's judgements. Using the criterion of European/non-European to predict White or non-White ethnic origin, there is agreement in 1876 (92.4%) of ICM's judgements and in 1868 (92.0%) of DNJL's judgements.

TABLE IV. Relationship between self-classified ethnic origin and origin of surname as assessed by ICM and DNJL for 2031 applicants to St Mary's. Values in italics represent those that are regarded as 'correct'. The vertical line divides candidates into those with European and non-European surnames, and the horizontal line into those defined as White or non-White.

Ethnic origin	Surname						Percentage
	British	European	Arab	Indian	Far Eastern	African	
<i>ICM</i>							
White, European	<i>1349</i>	<i>64</i>	4	26	9	5	71.7
Arab/Middle Eastern	2	4	<i>17</i>	<i>24</i>	2	0	2.4
Asian	39	16	<i>69</i>	<i>214</i>	10	14	17.8
Far Eastern	7	0	1	2	<i>66</i>	1	3.8
African	12	0	3	6	4	<i>11</i>	1.8
Caribbean	14	0	0	1	0	0	0.7
Other	13	4	4	12	0	2	1.7
<i>DNJL</i>							
White, European	<i>1334</i>	<i>79</i>	6	18	7	13	71.7
Arab/Middle Eastern	2	6	<i>15</i>	18	0	8	2.4
Asian	25	31	<i>52</i>	<i>228</i>	5	21	17.8
Far Eastern	9	1	1	1	<i>64</i>	1	3.8
African	12	4	0	6	2	<i>12</i>	1.8
Caribbean	13	1	0	1	0	0	0.7
Other	11	4	2	14	0	4	1.7

The sensitivity and specificity of non-European surnames for identifying non-White applicants can be analysed using standard epidemiological techniques (Rose & Barker, 1986). An ideal test should have a high sensitivity (i.e. it should detect a high proportion of those in the population who are actually non-White), a high specificity (i.e. a low proportion of Whites should be identified as non-White) and a high predictive value (i.e. a high proportion of those with non-European surnames will actually be non-White). These measures are relatively independent, and a test may perform poorly on one or several of the criteria.

A non-European surname is highly specific for identifying non-Whites (96.7%), only 3.3% of Whites having non-European surnames. There is, however, a somewhat lower sensitivity (84.6%), indicating that 15.4% of non-Whites have European surnames. The predictive value of a non-European surname is 88.3%, indicating that 11.7% of those with non-European surnames are actually White.

Validity of place of birth for identifying ethnic origin. Table V shows the relationship, in applicants who are UK nationals, between ethnic origin and place of birth, using either of two criteria, non-UK birth or New Commonwealth birth. For non-UK birth, the specificity is 95.2%, the sensitivity 38.9% and the predictive value 70.2% for identifying non-White applicants; for New Commonwealth birth the specificity is 98.1%, the sensitivity 33.3% and the predictive validity 84.0%. Neither measure is a good predictor of being ethnically non-White, principally because of a low sensitivity in each case, despite a very high specificity.

Analysis of residential neighbourhoods of applicants using post-codes. Neighbourhoods, as assessed from post-code analysis and ACORN coding, were analysed separately for White and non-White applicants (Tables VI and VII). It can be seen from Table VI that census variables provide a potential method for differentiating White and non-White applicants on almost all of the variables measured. Applicants to medical school generally come from more advantaged neighbourhoods than does the population in general, White applicants having more advantaged neighbourhoods than non-White applicants. ACORN coding of neighbourhoods (Table VII) shows that non-White applicants tend to live in neighbourhoods which are of lower status. In addition, most applicants from areas which have a particularly high number of non-White residents (H 26-29) are themselves non-White. Nevertheless the majority of non-White applicants do not live in such areas, and therefore ACORN codes will have a high specificity for identifying minority applicants but a low sensitivity. It is also of interest that, as might be expected, medical students in general come from higher status residential areas than does the population as a whole, although there are clear and quite numerous exceptions to that rule.

TABLE V. Relationship between place of birth and self-classified ethnic origin in UK nationals who applied to St Mary's

Ethnic group	N	Percentage not born in UK	Percentage born in New Commonwealth
White, European	1405	4.8	1.9
Arab/Middle Eastern	17	35.3	17.6
Indian subcontinent	301	41.5	36.2
African	21	14.3	14.3
Caribbean	10	20.0	10.0
Far Eastern	34	55.9	50.0
Other	28	17.9	14.3

Validity of neighbourhood data for identifying ethnic origin. Since neighbourhood data can differentiate between White and non-White applicants it remains to be seen whether such data can be used on their own, or as an adjunct to surnames, for identifying the ethnic origin of applicants.

TABLE VI. Comparison of small area census statistics based on the neighbourhoods in which applicants live, for White and non-White applicants (UK nationals only)

Neighbourhood characteristic	GB average 1981	White applicants		Non-White applicants		<i>t</i> -test
		Mean	SD	Mean	SD	
Unemployment (%)	9.0	5.9	4.3	8.7	6.3	<i>p</i> <0.001
Women in full-time work (%)	26.9	27.4	8.5	31.0	10.2	<i>p</i> <0.001
Students (%)	1.1	1.8	1.6	1.9	2.3	NS
One vehicle households (%)	45.5	47.5	10.1	44.0	11.1	<i>p</i> <0.001
2+ vehicle households (%)	15.9	25.0	14.8	17.1	13.5	<i>p</i> <0.001
4+ children households (%)	1.6	1.0	1.3	2.3	3.5	<i>p</i> <0.001
Afro-Caribbean born (%)	1.6	0.9	2.0	4.3	5.5	<i>p</i> <0.001
Indo-Pakistani born (%)	1.2	0.9	2.1	5.6	9.5	<i>p</i> <0.001
Age 0-4 (%)	6.0	5.2	2.4	6.1	3.1	<i>p</i> <0.001
Age 5-14 (%)	14.7	13.9	4.0	13.8	4.7	NS
Age 15-24 (%)	15.6	14.9	3.6	16.7	4.3	<i>p</i> <0.001
Age 25-44 (%)	26.3	26.3	6.3	27.1	6.3	<i>p</i> <0.025
Age 45-64 (%)	22.4	23.6	6.0	21.8	5.9	<i>p</i> <0.001
Age 65+ (%)	14.8	16.1	8.4	14.5	7.6	<i>p</i> <0.001
Population married (%)	49.8	50.6	6.3	47.3	7.1	<i>p</i> <0.001
Single non-pensioners (%)	7.1	7.6	6.8	9.6	7.5	<i>p</i> <0.001
Persons per household	2.8	2.7	0.4	2.8	0.5	<i>p</i> <0.002
Rooms per household	4.9	5.5	0.9	5.0	0.9	<i>p</i> <0.001
Owner-occupied (%)	56.2	71.2	25.5	63.3	29.4	<i>p</i> <0.001
Council and housing association (%)	33.4	15.2	23.3	21.2	29.8	<i>p</i> <0.001
Unfurnished rented (%)	7.8	9.4	10.7	9.4	9.8	NS
Furnished rented (%)	2.5	4.2	8.1	6.1	9.0	<i>p</i> <0.001
1.5+ persons per room (%)	0.9	0.5	1.5	1.5	2.7	<i>p</i> <0.001
1-1.5 persons per room (%)	3.8	1.6	2.0	4.1	4.8	<i>p</i> <0.001
Rooms per person	1.8	2.0	0.3	1.8	0.3	<i>p</i> <0.001
Shared amenities (%)	1.2	4.2	6.9	6.4	7.8	<i>p</i> <0.001
No inside WC (%)	2.6	1.8	3.6	2.5	4.9	<i>p</i> <0.01
No bath (%)	1.8	1.4	2.8	2.1	4.2	<i>p</i> <0.005
7+ rooms per dwelling (%)	11.9	23.1	17.2	15.3	15.6	<i>p</i> <0.001
1 or 2 rooms per dwelling (%)	5.6	5.1	9.6	7.6	10.6	<i>p</i> <0.001
Walk to work (%)	15.7	12.4	12.4	13.9	13.0	<i>p</i> <0.10
Public transport to work (%)	22.0	17.1	15.4	30.2	19.9	<i>p</i> <0.001
Manufacturing workers (%)	27.9	20.8	13.7	26.3	16.3	<i>p</i> <0.001
Agricultural workers (%)	2.1	2.7	7.4	0.2	1.5	<i>p</i> <0.001
Service workers (%)	52.4	61.4	16.9	57.3	17.8	<i>p</i> <0.001
Professional workers (%)	16.1	25.8	16.4	20.0	16.2	<i>p</i> <0.001
Non-manual workers (%)	31.4	37.0	14.7	36.1	17.1	NS
Skilled manual workers (%)	25.2	18.4	13.4	20.1	11.4	<i>p</i> <0.025
Semi-skilled workers (%)	19.3	13.6	11.2	17.3	13.8	<i>p</i> <0.001
Unskilled workers (%)	6.2	3.5	5.5	4.8	6.6	<i>p</i> <0.001
<i>N</i>			1332		397	

Note: Census data are derived from the 1981 census of Great Britain, and the GB average 1981 column shows the mean value for all small areas in that census.

Hierarchical discriminant analysis was used to estimate a linear discriminant function for distinguishing White and non-White applicants using the 40 neighbourhood variables described in Table VII, and the function was used to predict group membership, using population estimates as prior probabilities. Although a highly significant discriminant

TABLE VII. ACORN summary groups (A) and detailed types (B) for coding of neighbourhoods derived from the post-codes of White and non-White applicants compared with known proportions for the British population as a whole ('1981 population'). Percentage of all applicants gives the percentage of applicants in each of the neighbourhood types for the 1729 applicants for whom both ethnic group and ACORN code were known. Percentage of non-White applicants gives similar figures for the 397 non-White applicants, and percentage non-White gives the percentage of individuals in each neighbourhood group who are non-White.

		1981 Population	Percentage of all applicants ^a	Percentage of non-White applicants ^b	Percentage of non-White ^c	
(A) ACORN summary groups						
A	Agricultural areas	3.4	3.1	0.3	1.9	
B	Modern family housing, higher incomes	16.2	14.8	10.3	16.0	
C	Older housing of intermediate status	17.6	13.4	8.1	13.8	
D	Poor quality older terraced housing	4.3	1.5	2.0	30.8	
E	Better-off council estates	13.0	5.7	5.8	23.2	
F	Less well-off council estates	9.4	2.6	4.8	42.2	
G	Poorest council estates	7.6	1.6	3.8	53.6	
H	Multiracial areas	3.9	6.9	25.7	85.7	
I	High status non-family areas	4.2	11.3	12.3	25.0	
J	Affluent suburban housing	15.9	32.8	22.7	15.9	
K	Better-off retirement areas	3.8	6.1	4.3	16.2	
U	Unclassified	0.7	—	—	—	
(B) ACORN types						
A	1	Agricultural villages	2.6	2.5	0.3	2.3
	2	Areas of farms and smallholdings	0.8	0.6	0	0
B	3	Cheap modern private housing	4.1	2.3	1.0	10.0
	4	Recent private housing, young families	3.1	2.1	2.3	24.3
	5	Modern private housing, older children	5.8	5.5	4.8	20.0
	6	New detached houses, young families	2.6	4.7	0.5	2.3
	7	Military bases	0.5	0.2	0	0
C	8	Mixed owner-occupied and council estates	3.5	2.5	1.3	11.4
	9	Small town centres and flats above shops	4.0	3.9	2.8	16.4
	10	Villages with non-farm employment	4.6	4.3	2.0	10.8
	11	Older private housing, skilled workers	5.5	2.7	2.0	17.0
D	12	Unimproved terraces with old people	2.5	1.1	1.8	36.8
	13	Pre-1914 terraces, low income families	1.4	0.3	0	0
	14	Tenement flats, lacking amenities	0.4	0.1	0.1	30.0
E	15	Council estates, well-off older workers	3.6	2.1	1.8	18.9
	16	Recent council estates	2.6	1.2	2.8	55.0
	17	Council estates, well-off young workers	4.9	2.4	1.3	11.9
	18	Small council estates, often Scottish	2.0	0.1	0	0
F	19	Low rise estates in industrial towns	4.7	1.2	2.0	38.1
	20	Inter-war council estates, older people	3.1	1.0	1.8	38.9
	21	Council housing for the elderly	1.5	0.3	1.0	66.7
G	22	New council estates in inner cities	2.0	1.0	2.8	61.1
	23	Overspill estates, high unemployment	3.2	0.5	1.0	44.4
	24	Council estates with overcrowding	1.6	0	0	0
	25	Council estates with worst poverty	0.7	0.1	0	0
H	26	Multi-occupied terraces, poor Asians	0.4	0.7	3.0	100.0
	27	Owner-occupied terraces with Asians	1.1	1.9	7.1	87.5
	28	Multi-let housing with Afro-Caribbeans	0.7	1.0	4.3	94.4
	29	Better-off multi-ethnic areas	1.7	3.3	11.3	78.9
I	30	High status areas, few children	2.1	6.9	6.0	20.0
	31	Multi-let big old houses and flats	1.5	3.2	4.8	34.5
	32	Furnished flats, mostly single people	0.6	1.2	1.5	28.6
J	33	Inter-war semis, white collar workers	5.7	5.5	7.1	29.5
	34	Spacious inter-war semis, big gardens	5.0	11.7	7.3	14.3
	35	Villages with wealthy older commuters	2.9	5.3	1.8	7.6
	36	Detached houses, exclusive suburbs	2.3	10.2	6.5	14.7
K	37	Private house, well-off elderly	2.2	3.6	1.0	6.5
	38	Private flats with single pensioners	1.6	2.5	3.3	30.2
U	39	Unclassified	0.7	—	—	—
Total		100	100	100	—	

^a Percentage of applicants in each of the neighbourhood types for the 1729 applicants for whom both ethnic group and ACORN code were known.

^b Percentage of non-White applicants in each of the neighbourhood types for the 397 non-White applicants.

^c Percentage of individuals in each neighbourhood group who are non-White.

function was produced for discriminating White from non-White applicants ($p < 0.001$),

analysis of the sensitivity, specificity and predictive value of the discriminant function produced a less satisfactory picture, shown in Table VIII(A). Whatever items were included in the analysis, there was a low sensitivity and a low predictive value, despite a high specificity. That is, although most Whites were correctly identified as such, the method misclassified a high proportion of non-Whites as White. Successive addition of extra census variables did not help this problem.

Although not satisfactory when used on their own for predicting ethnic origin, it is possible that the neighbourhood variables might be more useful when combined with surnames. Table VIII(B) shows that in fact they produce no useful additional identification over and above that provided by surname alone. A final possible use of the neighbourhood

TABLE VIII. Sensitivity, specificity, predictive value and total correct predictions for linear discriminant functions for predicting the ethnic group of British medical school applicants based on surname, place of birth and 40 small area census statistics extracted using the applicants' post-codes

Step model	Sensitivity (%)	Specificity (%)	Predictive value (%)	Total correct (%)
(A) Census variables: stepwise analysis: all applicants				
1 Afro-Caribbean born	24.7	97.7	76.6	81.0
2 + Indo-Pakistani born	29.2	97.4	77.3	81.8
3 + Public transport to work	31.0	96.9	75.0	81.8
4 + 7+ rooms per dwelling	32.2	97.0	76.2	82.1
5 + Agricultural workers	32.5	97.0	76.3	82.2
6 + 1.5+ persons per room	33.0	97.1	77.5	82.4
7 + Manufacturing workers	32.5	97.3	78.2	82.4
8 + Women in full-time work	33.2	97.4	79.0	82.7
9 + One vehicle households	34.0	97.4	79.4	82.8
10 + Rooms per person	34.8	97.3	79.3	82.9
11 + remaining 30 variables	38.0	96.8	78.2	83.3
(B) Surname and census variables: hierarchical analysis, all applicants				
1 Non-European surname	84.4	96.7	88.3	93.9
2 + first 10 predictors above	84.6	96.7	88.4	93.9
(C) Census variables: stepwise analysis, applicants with European surnames				
1 Afro-Caribbean born	24.6	98.0	36.6	94.7
2 + 1-1.5 persons per room	26.2	98.0	38.1	94.7
3 + Indo-Pakistani born	24.6	97.7	34.1	94.4
4 + Rooms per person	26.2	97.9	37.2	94.7
5 + Public transport to work	26.2	97.9	37.2	94.7
6 + 4+ children households	26.2	97.9	37.2	94.7
7 + Unemployment	26.2	98.0	38.1	94.7
8 + Council and housing association	24.6	97.9	35.7	94.6
9 + Owner-occupied	26.2	97.8	36.4	94.6
10 + Unskilled workers	26.2	97.9	37.2	94.7
(D) Place of birth, census variables and surname: hierarchical analysis, all applicants				
1 Not born in UK	38.9	95.2	70.2	82.4
2 + 40 census variables	64.5	98.2	84.0	83.5
3 + Non-European surname	84.9	96.7	88.5	93.4
(E) Place of birth, census variables and surname: hierarchical analysis, all applicants				
1 Born in New Commonwealth	33.3	98.1	84.0	83.5
2 + 40 census variables	59.9	93.9	74.6	86.1
3 + Non-European surname	84.9	96.7	88.5	94.0

data might be in helping classify accurately that small group of applicants who are non-White but who have European surnames. Table VIII(C) shows a hierarchical discriminant analysis in the case of the applicants with European surnames only, distinguishing White from non-White applicants. The neighbourhood data are indeed capable of detecting non-White applicants in this group, but show the same pattern as in the overall data—of high specificity but low sensitivity—making them generally unsatisfactory for the task.

Table VIII(D) shows a hierarchical discriminant analysis for predicting ethnic origin in all applicants, using firstly place of birth (UK or not UK), followed by the census variables and then by non-European surname (NES). As in previous analyses, it is only when NES is included that an adequate sensitivity is achieved. Table VIII(E) shows a similar analysis but using Born in New Commonwealth rather than Not born in UK as a variable: once again the sensitivity is low unless NES is included.

It is possible that there is some particular combination of the various predictor variables (surname, place of birth and ACORN group H) which allows an adequate distinction between White and non-White applicants. Table IX shows the sensitivity, specificity and predictive value for all two- and three-way combinations of the variables. The use of one or two other variables in combination with NES does improve sensitivity, but only, in each case, at the expense of decreasing specificity. The conclusion must be that no combination of variables is an improvement upon the use of NES alone.

Table X shows the actual number of individuals in each combination of surname, ACORN group and place of birth, and the percentage in each group that is non-White and their contribution to the total pool of non-White applicants. Variables AGH and BNWCM only really differentiate between White and non-White applicants within the group with non-European surnames. It is for that reason that none of the combinations of predictors used in Table IX are substantial improvements over NES alone.

Differential survey responses by minority and non-minority applicants. Considering just the 1877 UK applicants in the survey, 1742 (92.8%) returned our questionnaire, and for these individuals self-classifications of ethnic origins are known. For the remaining applicants there is information only about place of birth, surname and ACORN code. Of 1829 questionnaire returners, 21.7% had non-European surnames, compared with 32.4% of 142 non-returners ($\chi^2=8.64$, 1 df, $p<0.005$), meaning that non-returners were 1.75 times more likely to have non-European surnames. Similarly, 8.9% of 1829 returners were born in the New Commonwealth, compared with 15.5% of non-returners ($\chi^2=6.86$, 1 df, $p<0.01$), non-returners being 1.89 times more likely to have been born in the New Commonwealth. Of 1742 returners, 7.0% lived in ACORN group H (multiracial areas), compared with 13.3% of non-returners ($\chi^2=7.27$, 1 df, $p<0.01$), non-returners being 2.04 times more likely to live in ACORN group H. Taken together these figures mean that non-responders in our survey are about 1.9 times more likely to show characteristics which are specific to applicants from minority groups. Since, however, only 7.2% of applicants did not return our questionnaire, we can conclude that the true, unbiased estimate of the proportion of non-White UK applicants is about 6.5% higher than that found in our data, i.e. 24.2% rather than 22.7%.

Since our data suggest that non-White applicants are somewhat less likely to have responded to our questionnaire, we may also ask whether non-White responders took longer to reply to the questionnaire, or were more likely to reply only after receiving a reminder. Non-White applicants replied on average 26.2 days (SD=21.1) after we sent our questionnaire, compared with 24.8 days (SD=21.9) for White applicants ($t=1.1$, 1806 df, NS). Neither were there significant differences in the rate of replying between the six commonest ethnic groups ($F(5,1774)=0.87$, NS). Of 1398 White applicants, 25.3% returned their

TABLE IX. Sensitivity, specificity and predictive value of different combinations of surname, place of birth and ACORN code for detecting members of ethnic minorities using different combinations of criteria

Criterion	Sensitivity (%)	Specificity (%)	Predictive value (%)	Total correct (%)
Non-European surname (NES)	84.4	96.7	88.3	93.9
Born in UK (BUK)	38.9	95.2	70.2	82.4
Born in New Commonwealth (BNWCM)	33.3	98.1	84.0	83.5
ACORN group H (AGH)	24.8	98.8	85.7	82.0
NES or BUK	88.6	92.3	77.1	91.5
NES or AGH	87.6	95.9	86.1	94.0
NES or BNWCM	87.1	94.9	83.4	93.2
AGH or BUK	54.7	94.1	73.1	85.2
AGH or BNWCM	50.1	96.9	82.7	86.3
NES or AGH or BUK	91.0	91.6	76.0	91.5
NES or AGH or BNWCM	90.3	94.1	81.7	93.2

questionnaire only after receiving at least one reminder, compared with 26.1% of 410 non-White applicants ($\chi^2=0.10$, 1 df, NS). We conclude that non-White applicants responded as quickly to our questionnaire as did White applicants.

Discussion

The questionnaire described in the Appendix is an adequate method of obtaining self-descriptions of ethnic origin, subjects finding it acceptable and being willing to complete it, at least for research purposes. The questionnaire was written by ourselves due to the unavailability of any more acceptable research instrument, despite enquiries to the CRE, the former Greater London Council's Anti-Racist Programme and the University of Warwick's Centre for Research in Ethnic Relations. In general, instruments recommended by these groups failed for our purposes because of inappropriate ethnic classifications (e.g. 'Irish' or 'Spanish'), or because phrasing did not seem satisfactory for the highly intelligent group that we were assessing. In order to ensure comparability between studies we suggest that the precise wording of our questionnaire is used in future studies, although we would be surprised if slight changes produced substantial differences in response rates or proportions of ethnic applicants. Nevertheless there is some slight evidence that minority applicants are less likely to reply to a questionnaire in which there are questions concerning ethnic origins, although this difference does not substantially bias the population estimates derived from our figures, and we have no direct evidence that the difference is *because* of the ethnic questions.

If applicants have not been asked directly for self-classification of their ethnic origin, which is probably the most accurate and acceptable way of obtaining such information, then indirect measures must be used. Surnames can be *reliably* assessed as European or non-European, and they also show high *validity*, the measure being both specific and sensitive. Nevertheless there is a proportion of applicants, particularly of Far Eastern and Caribbean origin, who are not detected by this method, having typical European surnames. In the context of identifying members of the Asian community in a community medicine context, Nicoll *et al.* (1986) have also shown the high specificity, sensitivity and predictive value of surnames and also, particularly, of both first and second names.

TABLE X. The proportions of applicants showing each combination of surname, place of birth and ACORN code, who are non-White, and their contribution to the total group of non-White applicants

Group			N	N (non-white)	Percentage non-White ^b	Percentage of non-White applicants ^b
NES ^a	AGH ^a	BNWCM ^a				
—	—	—	1362	40	2.9	9.7
—	—	+	25	13	52.0	3.2
—	+	—	36	11	30.6	2.7
—	+	+	0	0	—	—
+	—	—	205	165	80.5	40.1
+	—	+	61	56	91.8	13.6
+	+	—	94	93	98.9	22.6
+	+	+	33	33	100.0	8.0

^a For definitions, see Table IX.

^b For definitions, see Table VII.

The assessment of candidates' residential neighbourhoods by the use of post-codes and ACORN coding provides good measures which show that non-White applicants come from neighbourhoods which are less affluent than White applicants, and that certain ACORN residential codes indeed have a high proportion of minority residents. Nevertheless these differences, although highly reliable statistically, are not satisfactory for identifying minority applicants, due to a lack of sensitivity. A similar consideration applies to the use of place of birth for identifying the ethnic origins of applicants.

General Discussion

These studies have shown that applicants are willing to provide descriptions of their ethnic origins using a straightforward easily understood questionnaire, as long as they are assured that information is being collected to monitor and ensure equal opportunities, rather than as a potential instrument of discrimination in its own right. In July 1988, UCCA announced, on the basis of a decision by the CVCP, that

applicants will be invited to describe their ethnic background on their application form. Doing so will be voluntary, and this information will not be made available to admission staff in the universities until after admissions have been completed. (Anonymous, 1988)

In April 1989 the Secretary of State for Education and Science confirmed in a Parliamentary written answer that such monitoring would be carried out by both UCCA and the PCAS for applicants for admission in autumn 1990 (Hansard, 25 April 1989). From our work we feel sure that such a procedure will be acceptable to candidates if the UCCA and PCAS forms are modified to have a separate tear-off portion at one side asking questions broadly similar to those described here, the portion being detached before the form is copied and sent to universities, and if it is made clear that the information is collected principally for the statistical monitoring of equal opportunities. We believe that candidates, and probably also universities, would welcome the introduction of such a change, and we note that it is the stated policy of the Association of University Teachers (Cottrell, 1986).

It is also apparent from our work that a candidate's surname is a reliable and valid indicator of ethnic origin and hence that it could itself be used as an instrument of discrimination. Since it is the only information on the UCCA form which has such high validity, it may be necessary to consider further modification of the UCCA form so that the

detachable portion also contains the applicant's name, the form then being identified only by its serial number during the initial stages of selection. For applicants called to interview, or given an offer without interview, UCCA could provide a name via its central computer facility, a note being taken that the name has been requested. A similar consideration might also apply to place of birth.

Since anonymous application forms would inevitably produce inconvenience and a slowing of the selection process, we suggest that only if UCCA's monitoring finds clear evidence of substantial discrimination during selection should names also be removed from forms, thereby removing the principal source of information upon which discrimination could be based.

Acknowledgements

We are grateful to the many medical school applicants who took the trouble to complete our long questionnaires. We also thank the ESRC for providing research funding, the CRE for financial assistance in ACORN coding of post-codes, Ms Denise Hale of CACI for her assistance in deriving ACORN codes, Dr Diana N. J. Lockwood for her assistance in classifying the ethnic origin of surnames, and Dr Peter Sanders and Mr Andrew Dorn of the CRE for helpful discussions.

Correspondence: Dr I. C. McManus, Senior Lecturer in Psychology applied to Medicine, St Mary's Hospital Medical School, Imperial College of Science, Technology and Medicine, Norfolk Place, London W2 1PG, United Kingdom.

REFERENCES

- ANONYMOUS (1988) Antiracism, *British Medical Journal*, 297, p. 143.
- CACI (1985) *ACORN: user's guide* (London, CACI Market Analysis).
- COLLIER, J. & BURKE, A. (1986) Racial and sexual discrimination in the selection of students for London medical schools, *Medical Education*, 20, pp. 86-90.
- CRE (1984) *Monitoring an Equal Opportunity Policy: a guide for employers*, 3rd edn (London, CRE).
- CRE (1985) *Ethnic Minorities in Britain: statistical information on the pattern of settlement* (London, CRE).
- CRE (1988) *Medical School Admissions: report of a formal investigation into St George's Hospital Medical School* (London, CRE).
- COTTRELL, P. (1986) Policy for ethnic minorities: equal opportunity practice in universities, *AUT Bulletin*, November, p. 7.
- LOWRY, S. & MACPHERSON, G. (1988) A blot on the profession, *British Medical Journal*, 296, pp. 657-658.
- McMANUS, I.C. & RICHARDS, P. (1985) Admission to medical school, *British Medical Journal*, 290, pp. 319-320.
- MORGAN, M. (1983) Measuring social inequality: occupational classifications and their alternatives, *Community Medicine*, 5, pp. 116-124.
- MORGAN, M. & CHINN, S. (1983) ACORN group, social class, and child health, *Journal of Epidemiology and Community Health*, 3, pp. 196-203.
- NICOLL, A., BASSETT, K. & ULJASZEK, S.J. (1986) What's in a name? Accuracy of using surnames and forenames in ascribing Asian ethnic identity in English populations, *Journal of Epidemiology and Community Health*, 40, pp. 364-368.
- PAREKH, B. (1987) Higher education and racial discrimination, *AUT Bulletin*, June, pp. 6-7.
- ROSE, G. & BARKER, D.J.P. (1986) *Epidemiology for the Uninitiated*, 2nd edn (London, British Medical Association).
- SMITH, R. (1987) Prejudice against doctors and students from ethnic minorities, *British Medical Journal*, 294, pp. 328-329.
- SPELLER, V. & HALE, D. (1985) Making the most of your postcode, *Health and Social Service Journal*, 28 February, pp. 252-253.
- VEITCH, A. (1984) Has discrimination been diagnosed in the medical schools?, *The Guardian*, 8 August, p. 13.

Appendix. The Questionnaire Used by Applicants to Describe Their Ethnic Origins

Ethnic Origins

It is important that we are able to monitor the ethnic origins of applicants to medical schools, so that we can assess claims that equal opportunities might not apply to all applicants. We realise that such questions may be construed as controversial. Nevertheless please try and answer the questions as far as is possible, bearing in mind the reasons for the study, and that strict confidentiality will be maintained.

Please note. In assessing ethnic group we are *not* asking about nationality (which has been mentioned earlier), but rather we are interested in whether your physical appearance is similar to that of the majority of persons from particular parts of the world.

—If you were forced to classify yourself as one of the following ethnic groups, based primarily on your physical appearance, which would it be:

- | | | |
|--|---------------------|--------------------|
| White, European | Arab/Middle Eastern | Asian (i.e. Indian |
| African | Caribbean | subcontinent) |
| Far Eastern (i.e. China/Japan/East Indies, etc.) | | |
| Other (please specify) _____ | | |

—If your parents are not both in the same ethnic group as yourself, in which of the above groups would you place them?

- Father _____
 Mother _____

—Were your parents *born* in the United Kingdom?

- Neither/Father only/Mother only/Both

—How many of your grandparents were *born* in the United Kingdom?

- None/One/Two/Three/Four

—Is English your mother tongue? No/Yes

- If *No*, at what age did you first start to speak English? _____

—If you felt unable to answer any of the above questions, please try and give brief reasons for or comments about your difficulty.

—Are you in favour of attempting to monitor the ethnic composition of medical school applicants for the purposes of assessing quality of opportunity?

- Definitely not/Probably not/Probably yes/Definitely yes