

Ethnicity and academic performance in UK trained doctors and medical students: systematic review and meta-analysis

Katherine Woolf, lecturer in medical education,¹ Henry W W Potts, senior lecturer in health informatics,² I C McManus, professor of psychology and medical education¹

¹Academic Centre for Medical Education, UCL Division of Medical Education, London N19 5LW, UK

²Centre for Health Informatics and Multiprofessional Education, UCL Division of Population Health, London N19 5LW

Correspondence to: K Woolf
k.woolf@medsch.ucl.ac.uk

Cite this as: *BMJ* 2011;342:d901
doi:10.1136/bmj.d901

ABSTRACT

Objective To determine whether the ethnicity of UK trained doctors and medical students is related to their academic performance.

Design Systematic review and meta-analysis.

Data sources Online databases PubMed, Scopus, and ERIC; Google and Google Scholar; personal knowledge; backwards and forwards citations; specific searches of medical education journals and medical education conference abstracts.

Study selection The included quantitative reports measured the performance of medical students or UK trained doctors from different ethnic groups in undergraduate or postgraduate assessments. Exclusions were non-UK assessments, only non-UK trained candidates, only self reported assessment data, only dropouts or another non-academic variable, obvious sampling bias, or insufficient details of ethnicity or outcomes.

Results 23 reports comparing the academic performance of medical students and doctors from different ethnic groups were included. Meta-analyses of effects from 22 reports (n=23 742) indicated candidates of “non-white” ethnicity underperformed compared with white candidates (Cohen's $d=-0.42$, 95% confidence interval -0.50 to -0.34 ; $P<0.001$). Effects in the same direction and of similar magnitude were found in meta-analyses of undergraduate assessments only, postgraduate assessments only, machine marked written assessments only, practical clinical assessments only, assessments with pass/fail outcomes only, assessments with continuous outcomes only, and in a meta-analysis of white v Asian candidates only. Heterogeneity was present in all meta-analyses.

Conclusion Ethnic differences in academic performance are widespread across different medical schools, different types of exam, and in undergraduates and postgraduates. They have persisted for many years and cannot be dismissed as atypical or local problems. We need to recognise this as an issue that probably affects all of UK medical and higher education. More detailed information to track the problem as well as further research into its causes is required. Such actions are necessary to ensure a fair and just method of training and of assessing current and future doctors.

INTRODUCTION

In 1995, a *BMJ* news article reported that all the students who failed clinical finals at the University of Manchester the previous year had been men with Asian names.¹ A systematic review of the predictors of medical school success published seven years later found that white ethnicity predicted good performance, but only one of the 14 included reports came from the United Kingdom.²

Two large studies of degree outcomes in UK higher education have since shown that, across all subjects, white students were more likely than students who categorised themselves as Asian, black, mixed, or Chinese/other to achieve first or upper second class degrees. Differences in attainment between Asian and black students mostly disappeared when socioeconomic status was taken into account. Even after adjustment for up to seven confounding variables, however, the white students still achieved higher degree classes than students from all the minority ethnic groups.³⁻⁵

Medicine was largely excluded from these studies because it is an unclassified degree (that is, students either pass or fail; they do not receive first, second, or third class degrees). There is therefore less certainty about ethnic differences in the attainment of UK medical students, who are particularly highly selected for academic excellence and often come from privileged socioeconomic backgrounds.⁶⁻⁸ In postgraduate terms, ethnic differences in the academic attainment of doctors have been explored mostly only in terms of country of primary medical qualification.^{9 10}

A third of all UK medical students are from minority ethnic groups, 1.6 times the proportion on other undergraduate courses,¹¹ with by far the largest minority ethnic group being the Indian group (11%), followed by the Pakistani group (5%) (table 1). In 2009, 36% of newly qualified doctors and 52% of all other hospital doctors working in the NHS were from minority ethnic groups.^{12 13} The UK's Race Relations Amendment Act 2000 places a duty on all public authorities, including universities and the National Health Service, to monitor admission and progress of students and the recruitment and career progression of staff by ethnic group to be able to address inequalities or disadvantage.^{8 14 15}

Table 1 | Details of students accepted to study medicine and dentistry in UK in 2009 (Universities and Colleges Admissions Services, 2009)

Ethnic group	No (%)
White	5519 (67)
Mixed:	
White and Asian	172 (2)
White and black African	26 (0.3)
White and black Caribbean	28 (0.3)
Other	91 (1)
Asian:	
Indian	917 (11)
Pakistani	408 (5)
Bangladeshi	69 (1)
Other	331 (4)
Black:	
African	186 (2)
Caribbean	31 (0.4)
Other	6 (0.1)
Chinese	186 (2)
Other ethnic background	168 (2)
Unknown/prefer not to say	116 (1)
Total	8254 (100)

We undertook a systematic review and meta-analysis of studies comparing the academic performance of UK trained doctors and medical students from white and minority ethnic, or non-white, groups.

METHODS

This section gives details of the protocol for the review.

Ethnicity

The concept of “ethnicity” is complex, politically charged, and context specific.¹⁶ As such, we explicitly state how we interpreted ethnicity and our subsequent choice of ethnicity variable. In defining ethnicity, we followed Senior and Bhopal, who wrote: “[ethnicity] implies one or more of the following: shared origins or social background; shared culture and traditions that are distinctive, maintained between generations, and lead to a sense of identity and group; and a common language or religious tradition.”¹⁷

The white/non-white comparison

As with all reviews, we were restricted in our analyses by the data collected and reported in the original studies. In particular, we were restricted in our comparisons between different ethnic groups because most of the literature compared white candidates with all other—that is, non-white—candidates. While putting all the minority ethnic groups into one category for comparison with a white group was obviously not ideal, we had two reasons for this approach.

Firstly, the white/non-white comparison was a pragmatic approach to the lack of data on ethnicity in most studies, often because the numbers of candidates from certain minority ethnic groups were too small to allow sensible statistical analysis. We therefore compared the

ethnic group that was typically the largest—the white group—with all other groups combined—the non-white group. When information on the largest group after the white group—the Asian group—was available, we also performed that comparison. To an extent, all ethnic categories are essentially pragmatic because they can never take into account all the subtle variations between groups of people (for example, while the English census categories distinguish between people with their recent origins in India, Pakistan, and Bangladesh, they do not distinguish between those speaking Punjabi, Saraiki, Sindhi, Pashto, Urdu, Balochi, Kashmiri, etc).¹⁸

Secondly, the white/non-white comparison is scientifically justified by the evidence from UK higher education, which shows that the largest and least explained gap in attainment is between the white and non-white groups. This suggests that the white/non-white distinction is important in examination of the possible causes for this gap. Underlining this, the white/non-white distinction also seems important in other areas—for example, a recent report from the UK Government’s Department of Work and Pensions showed that, on average, children from all minority groups had higher levels of poverty than children in the white majority group.¹⁹

Types of reports, participants, and outcome measures

We included all published and unpublished quantitative reports on the academic performance of UK trained medical students or doctors that included a measure of candidate ethnicity.

All studies in our review included medical students undertaking formative or summative assessments at UK medical schools and UK trained doctors undertaking formative or summative UK postgraduate medical assessments.

The outcome measure can be encapsulated as “academic performance.” This includes pass/fail, attainment of other academic related specific goals (such as achieving a placement or not), and mean assessment scores.

Main comparison groups

The comparison was between white and non-white candidates. Where the data were available, we also compared white and Asian (Indian, Pakistani, and Bangladeshi) candidates.

Other important factors

We chose, a priori, to conduct separate meta-analyses for postgraduate and undergraduate assessments; machine marked written and practical clinical assessments; and pass/fail outcomes.

Exclusion criteria

We excluded reports with data only from outside the UK; self reported examination performance; lack of information about ethnic group (country of primary medical qualification for doctors and fee status for

undergraduates were insufficient); obvious sampling bias; lack of sufficient detail from which we could calculate an effect size and standard error; and outcome measures unrelated to academic attainment or that could be influenced solely by a non-academic factor (such as dropout).

Search strategy for identification of reports

KW had recently completed a PhD on ethnic differences in medical school performance,²⁰ and we used her personal knowledge of reports as a starting point.²¹ We also included data from other projects the authors were working on or were asked to advise on. We searched the online databases PubMed, Scopus, and ERIC using the following search terms:

- PubMed: (ethnic* OR race OR minority OR Asian) AND (“Education, Medical”[Mesh] OR “Educational Measurement”[Mesh]) AND “Great Britain”[Mesh]
- Scopus: ((undergraduate OR postgraduate) AND (“medical education”) AND ((perform* OR assess* OR exam* OR score OR grade OR fail)) AND (AFFILCOUNTRY((“great britain” OR “united kingdom” OR “northern ireland” OR england OR scotland OR wales)) AND ((ethnic* OR race OR minority OR asian)))
- ERIC: ((Keywords: ethnic) or (Keywords: minority) and (Thesaurus Descriptors: “Physicians” OR Thesaurus Descriptors: “Medical Education” OR Thesaurus Descriptors: “Medical Students” OR Thesaurus Descriptors: “Medical Schools”) not (Keywords: american) not (Keywords: states) not (Keywords: gpa) not (Keywords: MCAT)

We conducted specific searches using the search terms (ethnic* OR race OR Asian OR minority) of the e-journal versions of *Medical Education* (1966-2010), *Medical Teacher* (1979-2010), *Advances in Health Sciences Education* (all volumes), and *BMC Medical Education* (all volumes), as well as available published abstracts of conference proceedings of the Annual Scientific Meeting of ASME (Association for the Study of Medical Education) and the AMEE (Association for Medical Education Europe) annual conferences. We also used Google Scholar and Google to search the grey literature for government reports, etc. Finally, we used backwards and forwards citation searching.

Methods of the review

KW assessed reports for eligibility against previously agreed criteria and for methodological quality without consideration of their results. HWWP assessed those chosen reports. Reports were not assessed blind; we knew the authors' names, affiliations, and the source of publication. We discussed any differences until these were resolved. KW and ICM extracted data from full text versions of all included papers. All authors double extracted data from a sample of randomly chosen sources and reconciled any differences.

When reports had insufficient data for analysis, we contacted authors to ask for more complete data.

Synthesis of results and statistical analysis

To combine reports, we calculated an effect size (Cohen's *d*) and standard error for each.²² For categorical outcome variables, we first calculated an odds ratio and its associated confidence interval (www.hutchon.net/ConfidOR.htm) and then followed Chinn's method²³ to convert these into Cohen's *d* and standard errors.

When reports contained data from assessments taken at different points of the course by the same participants, we prioritised finals (year 5 assessments) over other undergraduate data as those examinations determine whether a medical student can become a doctor. Otherwise, we chose those taken by a larger number of candidates (larger sample size). When reports contained continuous measures of performance (such as exam score) and pass/fail data for the same examinations taken by the same participants, we prioritised continuous data because they are more sensitive; published categorical data, however, took precedence over unpublished continuous data.

When participants took more than one assessment in the same year, we calculated a mean score and standard deviation for all assessments, from which we calculated an effect size. When reports contained multivariate analyses, we prioritised simple effects; however if no simple effects were reported, we included outcome measures adjusted for other variables. In addition, we did a narrative summary of the effects of ethnicity adjusted for other variables.

When outcomes for separate minority ethnic groups were given, we back calculated the numbers who passed or the score for each group and combined groups as necessary to create a non-white category and to make the results more comparable with other reports. We also conducted a separate meta-analysis for white and Asian candidates. We used the definition of Asian given in the reports. When we had raw data, we defined Asian as Indian, Pakistani, or Bangladeshi (census categories).

Whenever possible, we analysed the performance of undergraduates with “home” (UK) or EU status only. In postgraduate examinations, we analysed the performance of UK graduates only.

Using MIX software (www.mix-for-meta-analysis.info/), we performed eight meta-analyses, one on all reports and seven on subsets of the data. We used random effects models and drew funnel plots for each to assess publication bias.

RESULTS

Searches

Figure 1 shows the number of reports, their identified sources, and reasons for exclusions.

Before the start of searching, we knew of 26 reports to include. Of these, 18 were published in peer reviewed journals,¹²⁴⁻⁴¹ two were KW's PhD²⁰ and ICM's unpublished data that supplemented the data

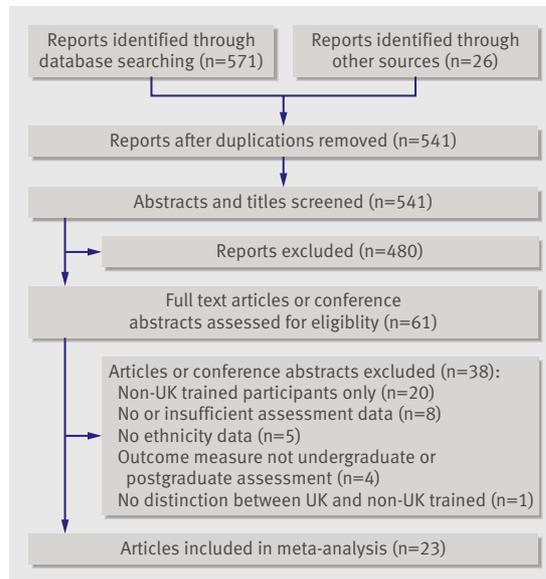


Fig 1 | Flowchart showing reports retrieved, excluded, and articles included in review

published in the study by McManus and Richards²⁴ and McManus and colleagues,²⁸ and one was published in the grey literature.⁹ The five remaining were retrieved from collaborations the authors were involved in around the time the review took place. These were Carroll and Mackenzie's conference abstract,⁴² supplementary unpublished data (M Carroll, personal communication, 2010), data from a pilot assessment for selection into specialty training in England run by the Association for Medical Royal Colleges (AoMRC, unpublished, 2010), a conference poster,⁴³ and an interim report of 2008 nMRCGP (new Membership of the Royal College of General Practitioners) exam results.⁴⁴

We retrieved 571 reports, including Brown⁴⁵, Ricketts⁴⁶, and Calvert⁴⁷ via online database searching. We found one report when we searched conference abstracts⁴⁸ and one when we searched the journal *Medical Teacher*.⁴⁹ We removed 56 duplicates, reviewed the abstracts and titles of the 541 remaining, and excluded 480. The full text versions of 61 reports were reviewed and another 38 excluded (20 did not provide data on UK trained participants; eight lacked sufficient details about outcomes; five lacked sufficient detail about participants' ethnic group; four did not have an appropriate assessment outcome; and one did not distinguish between UK and non-UK trained doctors).

We eventually included 23 reports in the meta-analysis. Many reports contained more than one set of data, for either the same or different candidates. Details of the studies, including factors on which quality was assessed, are given in table 2 (undergraduate prospective studies), table 3 (undergraduate retrospective studies), and table 4 (postgraduate). McManus (I C McManus, personal communication) provided the data referred to in the study by McManus and

Richards.²⁴ This study was excluded because it referred only in the text to an analysis by ethnicity but did not provide data on that analysis. McManus also provided supplementary data for another published study (McManus et al²⁸).

Undergraduate reports

Sixteen reports measured academic performance in undergraduates.

Design and sampling

One study was a cluster randomised controlled trial with ethnicity as one of the independent variables.³⁹ All others were prospective or retrospective cohort studies. Most studies combined more than one cohort of students from the same medical school. The largest study had over 2000 participants³⁰ and the smallest had 164.³⁸ All studies in which the outcome measure was a continuous measure of exam score had to exclude participants without exam data. This was typically about 5% of students. Five studies gave reasons for candidates lacking exam data.^{30 32 36 38 40} Ricketts et al included only students who progressed normally throughout the course, excluding those who re-sat exams or dropped out.⁴⁶

Fee status

Overseas students are likely to be educationally different from UK students.²⁸ Most reports differentiated between home students and overseas students, and two also adjusted for fee status in multivariate analyses.^{38 40} Lumb and Vai³² and Haq et al³³ looked only at UK students' performance. McManus et al performed separate tests for UK only students.²⁸ Wass et al reported that only two students in their cohort were not educated in the UK.³¹ In the study by Yates and James, ethnicity was unknown for 96% of overseas students, so this group was largely excluded.³⁶ Kilminster et al reported 7% of overseas students in their sample and said there were no differences in results between home and overseas students.⁴¹ Six studies^{20 30 37 39 42 46} and two sets of unpublished data (M Carroll, personal communication, 2010; McManus unpublished) did not distinguish between UK and overseas students in their analysis. The proportions of overseas students in these samples are probably small because most UK medical schools are allowed to take only about 7.5% overseas students (www.medschools.ac.uk/Students/Pages/FAQs.aspx#section8).

Outcome measures

All of the studies used formal summative assessments as outcome measures. Most gave explanations of their examinations: the format, how they were marked, and the subject matter covered. Two studies gave details of how the validity of the assessments had been established.^{31 46} Five reported the psychometric reliability of their assessments.^{20 31 33 37 39} Ten published reports^{20 31 33 36 38 41 42 46} and Carroll (personal communication, 2010) included continuous measures of assessment, and five included examination failure

Table 2 | Summary of prospective studies* of undergraduate medical students included in meta-analysis. For reports giving data for more than one cohort separately, number of candidates shown as range

Author	Sampling and sample size	Outcomes	Ethnicity details	Assessment details given	Statistical analyses (α set at 5% unless otherwise stated)
McManus ²⁸	Two cohorts of applicants to St Mary's in 1981 and 1986 who took University of London medical school finals. UK candidates n=576. Of total cohort of 691 who had taken finals, ethnicity data missing for 20. No reported details of students who dropped out before finals	Finals: clinical, MCQ, essay, and oral assessment marks. Overall pass/fail. Failure of at least one clinical exam	1981: European and non-European surname. 1986: self reported ethnicity. Compared UK white with UK non-white	Yes	Univariate tests and multivariate tests adjusted for sex, O levels, A levels, previous medical school performance, intercalated BSc
McManus (unpublished)	Two cohorts of applicants to St Mary's in 1981 and 1986 who entered various UK medical schools and took preclinical assessments (candidates described in McManus ²⁶), n=579-854. One cohort of medical students who applied in 1991 to St Mary's Hospital Medical School, University College, and Middlesex School of Medicine (now UCL Medical School), United Medical and Dental Schools of Guy's and St Thomas' Hospitals, University of Sheffield Medical School, or University of Newcastle upon Tyne Medical School (n=2907) and took London medical school finals (see McManus ²⁹ for details of that cohort). Supplementary data for McManus ²⁸	Preclinical (years 1 and 2): pass/fail (fail, retake year, retake one exam, pass, merit, or distinction). Finals: pass/fail (fail, retake year, retake one exam, pass, merit, or distinction)	1981 and 1986: see McManus. ²⁸ 1991: ethnic origin as reported by candidate on application form, with supplementary information derived from questionnaire (see McManus ²⁹)	Preclinical 1981 and 1986 cohorts: yes. ²⁴ Finals 1991 cohort: no	Univariate tests
Lumb ³²	700 entrants to Leeds in 1994-7, registered as home students who took year 3 OSCE. 38 students dropped out or did not take year 3 assessments. 21/38 required to leave; 15 transferred; 1 left because of ill health, and 1 completed PhD	Year 3 OSCE score	UCAS forms. Compared white, non-white, and unknown (n=4)	No	Multiple regression adjusted for sex, socioeconomic group, GCSE, and A level points, whether interviewed or not, shortlisting score, school type, and whether applied straight from school. No α reported. Raw data supplied by authors, allowing for univariate analysis
Yates ³⁶	590 students from three consecutive cohorts of students entering Nottingham. Numbers reduced on later outcome variables because 19 left in first 2 preclinical years, 9 didn't complete clinical course. 34 students spent >5 years on course	>1 preclinical exams failed; top/bottom 15th centile years 1 and 2; class of BMedSci; >1 clinical exams failed; top/bottom 15th centile clinical exams and skills exams; BMBS award	UCAS form or college records, converted into white and non-white	Yes	Univariate tests and logistic regression analysis adjusted for GCSE and A level grades and subjects, sex, and late offer. Subgroup analyses within non-white group showed no consistent differences on univariate analysis and so white v non-white maintained for multivariate analyses. α not reported
Woolf ³⁹	One cohort taking year 3 assessments at UCL in 2007. 335/352 students: 13 lost to follow-up with 6 having no exam data (reasons unknown) and 7 missing ethnicity	Year 3 OSCE and MCQ scores	Medical school records. Non-white groups aggregated to compare white and non-white	Yes	Univariate tests and multivariate tests to adjust for intervention. α not reported
Yates ⁴⁰	One cohort of entrants to Nottingham in 2007 who had taken UKCAT (n=195/260). 86% UK students. Excluded: 46 without UKCAT, 10 refused to consent, 4 transferred out of medicine, 3 transferred within course to BSc, 2 without exams. Included and excluded groups not significantly different	Preclinical (years 1 and 2) scores	Medical school records. Non-white groups aggregated to compare white and non-white	Yes	Univariate tests. Multivariate tests adjusted for A level grades and subjects passed, UKCAT scores, sex, domicile (home or overseas), selective or non-selective schooling. Significance set at P<0.01

MCQ=multiple choice questions; OSCE=objective structured clinical examination; UCL=University College London; UKCAT is aptitude test used to select medical students.

*Study by Woolf³⁹ is randomised controlled trial, all others are cohort studies.

(M Carroll, personal communication, 2010; McManus, unpublished).^{30 36 42} Eleven included written outcomes of machine marked assessments (M Carroll, personal communication, 2010; McManus, unpublished).^{20 33 36-39 41 42 46} Eleven included outcomes of practical assessments (M Carroll, personal communication, 2010; McManus, unpublished).^{20 31-33 36-39 41} In the study by Yates and James, the outcome for which an ethnic difference was reported (theme C) was assessed with a combination of written, online, oral presentation, and coursework.⁴⁰

Statistical analyses

Most studies conducted multivariate statistical tests (that is, with more than one predictor) to adjust the outcomes for factors other than ethnicity. Three published studies^{31 33 42} and two unpublished (M Carroll, personal

communication, 2010; McManus, unpublished) conducted only univariate analyses (looking solely at the effect of ethnicity on outcomes), although Haq et al restricted his sample to English speaking Asian UK students.³³ Most of the studies considered a P<0.05 to be significant, though three used P<0.01.^{31 40 41} Yates et al used P<0.001 as significant for univariate tests but did not state the significance level for multivariate tests.³⁸

Postgraduate reports

Seven reports measured postgraduate performance (table 4).

Design and sampling

All postgraduate reports were retrospective cohort studies except for one, which was both prospective and

Table 3 | Summary of retrospective cohort studies measuring ethnic differences on performance in undergraduate medical assessments. For reports giving data for more than one cohort separately, number of candidates is shown as range

Author	Sampling and sample size	Outcomes	Ethnicity details	Assessment details given	Statistical analyses (α set at 5% unless otherwise stated)
James ³⁰	2270 entrants to Nottingham 1975-90. 4 five year cohorts, size range n=336-719. 148 dropped out (59 health reasons, 12 academic reasons, 77 "medicine was not course they wished to pursue")	Success (obtaining BMedSci, first class, obtaining BMBS, BMBS honours)	Unknown. Collapsed into white and non-white	No	Univariate tests and multivariate logistic regression adjusted for at least 13 background variables
Wass ³¹	175/179 students who took finals exams at Guy's, King's and St Thomas' in 1999. All but two schooled in UK. Excludes 4 with unknown ethnicity	Final year OSCE score	Unknown	Yes	Univariate tests at P<0.01
Yates ³⁸	Entrants to Nottingham in 2000 who progressed normally, n=164/210. Excluded: 25 graduated late because of health/academic problems, 12 voluntary withdrawals, 2 failed year 1, 7 unknown A levels or overseas qualifications. Exclusions and inclusions not statistically different	Preclinical knowledge, preclinical skills; part II; clinical knowledge; clinical skills	Medical school and university databases. Collapsed into white and non-white	Yes	Univariate tests at P<0.001. Logistic regression adjusted for fee status, sex, entry qualifications, previous medical school performance
Kilminster ⁴¹	Three cohorts from different medical schools (n=709; cohort sizes n=209-353) who took year 3 assessments in 2002. Excludes 3 with missing ethnicity	Year 3 OSCE and written scores	Self report questionnaire. Also asked whether UK and whether English first language	Yes	Univariate tests at P<0.017. Multiple regression adjusted for sex at P<0.01
Haq ³³	1216 students in 2 consecutive cohorts at Imperial and UCL who took year 3 assessments in 2002 and 2003 (cohort sizes n=288-323). Asian and white UK students with English as first language	Year 3 MCQ and OSCE scores	Self reported questionnaire (50% response). Remainder: 2 researchers independently assigned to white or Asian with photos and by asking staff familiar with students; 95% inter-rater agreement. Differences resolved by discussion	Yes	Univariate tests
Woolf ³⁷	Two cohorts of students from UCL (n=363) and Imperial (n=331) taking year 3 assessments in 2004. Excludes 28 without exam data and 7 without ethnic data or photo. (Data from another UCL cohort previously reported in Haq ³³)	Year 3 OSCE and written multiple choice exam scores	Medical student records for 63%. Remainder categorised into white or non-white on basis of photo and name by 2 researchers independently; 3 differences resolved by discussion	Yes	Univariate tests. Analysis of covariance adjusted for sex and other exam performance
Woolf ²⁰	Two consecutive cohorts of UCL students taking year 3 assessments in 2005 (n=376) and 2006 (n=297). Excludes 3 students without ethnicity data	Year 3 OCSE, MCQ, and overall end of year scores	Medical school records. Collapsed into white and non-white	Yes	Univariate tests. Analysis of variance adjusted for sex
Ricketts ⁴⁶	10 cohorts of students over 5 years, all completing progress tests at Peninsula medical school in 2006-7 (n=746) and 2007-8 (n=819). Excludes those without complete assessment records	Aggregated multiple choice test score (progress test)	Unknown	Yes	Univariate tests. Multiple regression adjusted for year, disability, sex, and interaction terms. α not reported
Carroll ⁴²	One cohort who took year 1 and year 2 assessments in 2008 and 2009 at Bart's and the London (n=292)	Year 1 and year 2 written scores; pass/fail	Medical school records. Collapsed into white and non-white	Yes	Univariate tests
Carroll (2010, personal communication)	Three consecutive cohorts of students taking year 1 (n=283), year 2 (n=267), and year 3 (n=257) assessments in 2008 and 2009 at Bart's and the London	Year 1, year 2, and year 3 written scores. Year 3 clinical scores	Medical school records. Collapsed into white and non-white	No	Univariate tests

MCQ=multiple choice questions; OSCE=objective structured clinical examination; UCL=University College London.

retrospective.³⁴ The two largest studies had over 2000 candidates^{27,35}; the smallest had 53.⁴⁹ The studies were all cross sectional and therefore it was not appropriate to report the candidates lost to follow-up. Candidates were often excluded if ethnicity data were missing (though Dewhurst et al analysed the 10% missing ethnicity data separately³⁵). Two studies by Brown et al had some of the largest proportions of missing data, with 27% and 25% of their candidates being excluded because of missing data.^{45,49} However, this includes non-UK candidates whose data were not meta-analysed in this study.

Outcome measures

Binary outcome measures (pass rates or selection success) were included in all reports except for the report by Wakeford et al²⁷ and the unpublished 2010 report from the Association for Medical Royal Colleges, both of which used mean assessment scores. The two studies by Brown et al also included continuous measures of candidates' shortlisting, interview, or assessment scores.^{45,49} Three reports included data on written assessment performance,^{27,34,44} and four included data on practical assessment performance.^{27,34,35,44} All were formal summative assessments, except the

Table 4 | Summary of included reports measuring ethnic differences on performance in postgraduate assessments. All candidates were doctors

Author	Study design	Sampling and sample size	Outcome measures	Ethnicity details	Assessment details given	Statistical analyses
Wakeford ²⁷	Retrospective cohort study	Trainee GPs in 5 sittings of MRCGP 1988-90, n=3049-3326, excluding non-UK Asians	MCQ, MEQ, oral, aggregated essay paper and critical reading paperscores. Only those in top 85% of written exams go on to oral	Identified as Asian (Indian, Bangladeshi, Pakistani, or Sri Lankan) by first name and surname. Compared UK-trained Asians with non-Asians. Unknown whether non-Asians were born/trained in UK or not	Details of scoring and progression given. Details of marking given in reference to another paper	Univariate tests. Significance set at $\alpha=0.05$. First attempts only
Bessant ³⁴	Prospective and retrospective cohort study	483/534 of doctors on PACES course who subsequently responded to questionnaire and took next PACES, n=227 UK candidates; included 122 (50 UK) who had previously failed PACES	PACES pass rate	Ethnicity self reported in questionnaire. Compared UK white and UK non-white groups	No	Univariate tests. Logistic regression to adjust for passing part II written first time, age, sex, having p BSc, having been on previous PACES course, work experience. α not reported
Dewhurst ³⁵	Retrospective cohort study	UK trained candidates taking MRCGP(UK) parts 1, 2 and PACES in 2003-4, n=2528-5139. 10% ethnicity missing	Pass rate and raw scores for parts 1, 2 and PACES	Ethnicity self declared by questionnaire. Compared UK white with six other UK non-white categories and unknown group	Yes	Univariate tests. Logistic regression to adjust for sex and attempt number. Separate analysis for first attempt only. α not reported.
RCGP ⁴⁴	Retrospective cohort study	n=1036 UK white and Asian candidates taking nMRCGP exam in 2008-9. Includes resitters.	Clinical skills assessment and applied knowledge test pass rates	Unknown how ethnicity recorded. Compared UK white and UK Asian	No	Univariate tests. Adjusted figures not available. α not reported
Brown ⁴⁵	Retrospective cohort study	All candidates for GP training posts in West Midlands over four rounds October 2000 to March 2002. 27% of total sample excluded because ethnicity or country of qualification unknown or did not fit into category Not clear if duplicate candidates in different rounds, n=359 UK candidates (comprising 4 cohorts, n=54-125)	Proportion successful at shortlist and those placed. Candidates' shortlist and total scores (aggregated shortlist and assessment centre scores)	Ethnicity measured by application form and equal opportunities monitoring forms. Compared UK white and UK non-white groups	Yes	Univariate tests. Significance set at $\alpha=0.05$
Brown ⁴⁹	Retrospective cohort study	All candidates for GP training posts in West Midlands in 2000. 25% of total sample excluded because ethnicity or country of qualification unknown, n=53 UK candidates	Proportion successful at longlist, shortlist, interview, placed. Also candidates' scores at each of those stages	Ethnicity measured by application form and equal opportunities monitoring forms. Compared UK white and UK non-white groups	Yes	Univariate tests. α not reported
Association of Medical Royal Colleges (2010, unpublished)	Retrospective cohort study	n=462 UK trained volunteers taking pilot assessment for selection into specialty training in GP, ACCS, anaesthetics, paediatrics, CMT, histopathology, and psychiatry across 25 venues in England. Ethnicity missing for 1 candidate	Clinical problem solving test scores. Single best answer written test. Machine marked	Compared UK white with UK candidates from 6 non-white groups. Also compared UK white with UK non-white, and UK white with UK Asian	Not applicable, unpublished	Univariate tests. Multiple regression to adjust for sex, age, and preferred specialty. Significance set at $\alpha=0.05$

MCQ=multiple choice questions; MEQ=modified essay question; PACES=practical assessment of clinical examination skills; ACCS=acute care common stem; CMT=core medical training.

unpublished 2010 report from the Association for Medical Royal Colleges, which was a pilot test. Two studies restricted analyses to first attempts only.^{27 35}

Statistical analysis

Four studies reported only univariate analyses.^{27 44 45 49} The remainder used multivariate analyses to adjust for variables including sex, previous exam performance, and age. When it was reported, all studies considered $P<0.05$ as significant.

Meta-analysis 1: all reports

We included data on 23 742 candidates from 22 reports (36 datasets) in the meta-analysis of all reports. We excluded the second paper by Brown et al⁴⁹ as it contained data on the same candidates as in their other paper.⁴⁵ Overall, 17 172 candidates were white and

6570 non-white. The negative effect of non-white ethnicity on performance was significant ($P<0.001$) and of medium magnitude ($d=-0.42$; 95% confidence interval -0.49 to -0.34) (fig 2). A funnel plot showed no obvious publication bias (fig 3). There was heterogeneity in the sample ($I^2=72\%$). Of the 36 datasets, 35 showed a negative effect of non-white ethnicity and 25 of those showed a significantly negative effect. One showed no effect. None showed a positive effect of non-white ethnicity. The funnel plot showed no sign of publication bias.

Meta-analysis 2: undergraduate assessments

We included data on 13 193 undergraduates from 16 reports (27 datasets). The negative effect of non-white ethnicity on performance remained significant ($P<0.001$), with the same effect size ($d=-0.42$, -0.49

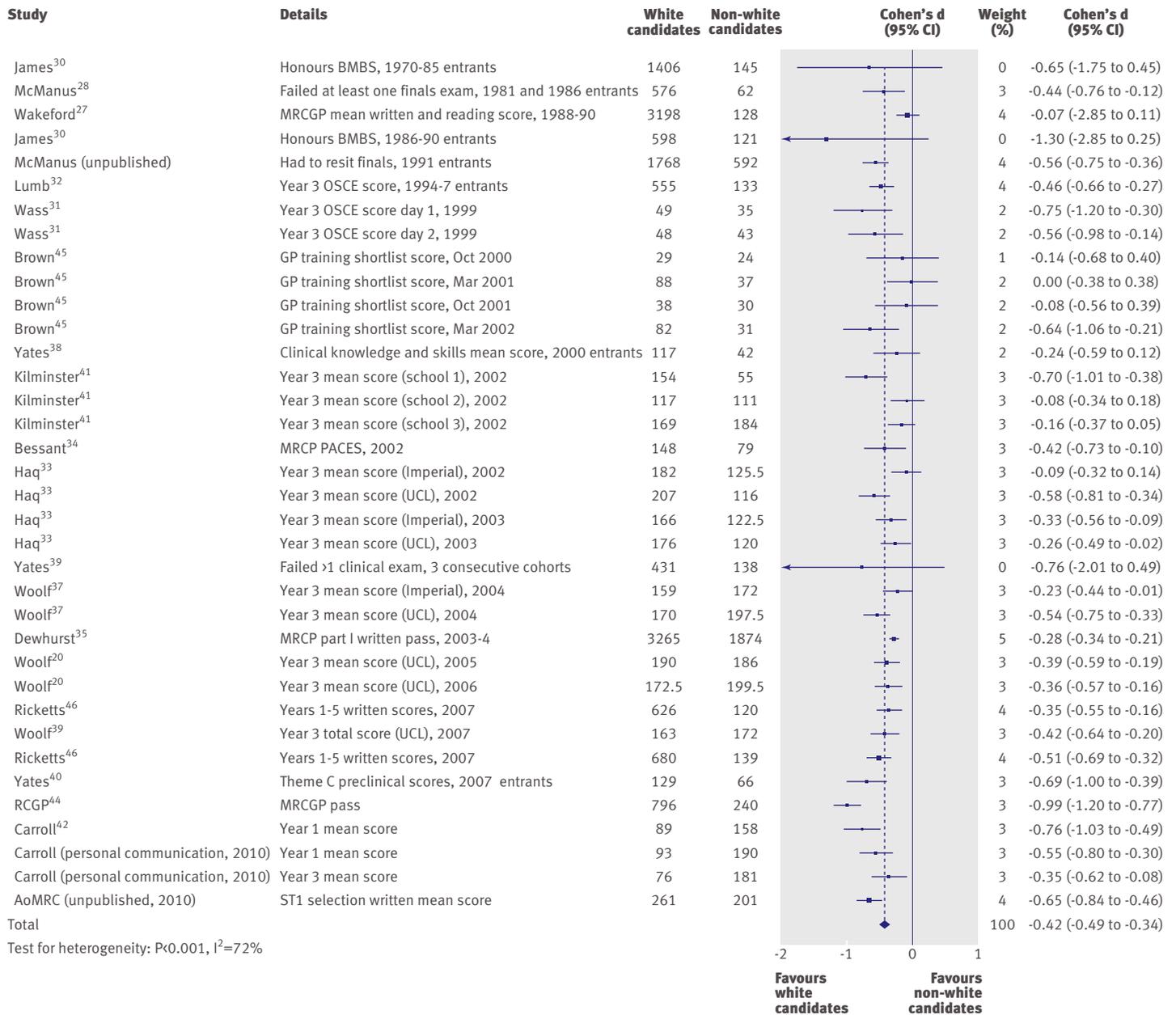


Fig 2 | Meta-analysis of all included studies of effect of non-white ethnicity on academic performance in medical students and postgraduates trained in UK. Negative signs are arbitrary and correspond to how data were coded, in this case white=0 and non-white=1; thus a negative sign indicates negative effect on non-white ethnicity on performance

to -0.35). There was less heterogeneity in this subset than in the meta-analysis of all results ($I^2=50\%$) (fig 4). The funnel plot showed no sign of publication bias.

Meta-analysis 3: postgraduate assessments

We included data on 10 549 postgraduate candidates from six reports (nine datasets). We excluded the 2001 paper by Brown et al⁴⁹ as it contained data on the same candidates as in their 2003 paper.⁴⁵ The negative effect of non-white ethnicity on performance remained significant ($P<0.001$), with a similar effect size ($d=-0.38$, -0.60 to -0.17). Heterogeneity was present ($I^2=89\%$) (fig 5). The funnel plot showed no sign of publication bias.

Meta analyses 4-7

In the meta-analysis of machine marked assessments, we included data on 20 415 candidates from 14 reports (26 datasets). The negative effect of non-white ethnicity on performance remained significant ($P<0.001$; $d=-0.35$, -0.44 to -0.26; $I^2=81\%$).

In the meta-analysis of practical assessments we included data on 16 038 candidates from 15 reports (27 datasets). The negative effect of non-white ethnicity on performance remained significant ($P<0.001$; $d=-0.42$, -0.52 to -0.33; $I^2=76\%$).

In the pass-fail meta-analysis we included data on 10 990 candidates from nine reports (10 datasets). As the paper by Carroll⁴² and the extra data provided by

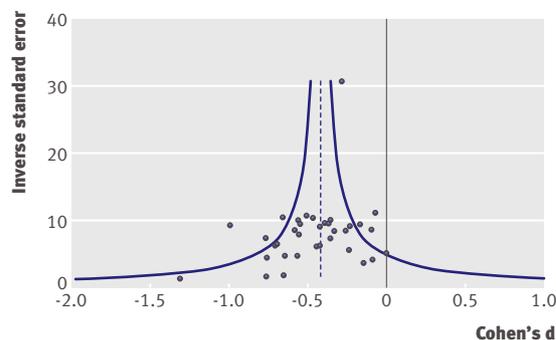


Fig 3 | Funnel plot of effect size by inverse standard error, showing no evidence of publication bias in studies of ethnicity and academic performance in trained doctors and medical students

Carroll (personal communication, 2010) included overlapping data we included only the former in the meta-analysis. We conducted two meta-analyses of pass/fail outcomes. The first used the odds ratios converted to effect sizes for inclusion in meta-analysis 1, and showed a significant negative effect of non-white ethnicity on performance ($P < 0.001$, $d = -0.59$, -0.84 to -0.35 ; $I^2 = 83\%$). The second used the original odds ratios reported in the studies, to check for bias in the conversion process, and again showed a significant negative effect of non-white ethnicity on performance ($P < 0.001$, odds ratio 2.92, 1.88 to 4.55; $I^2 = 83\%$).

In the continuous outcomes meta-analysis we included data on 12 174 candidates from 13 reports (30 datasets). The negative effect of non-white ethnicity on performance was significant ($P < 0.001$; $d = -0.38$, -0.46 to -0.30 ; $I^2 = 64\%$).

Where the data were available, we conducted a separate analysis comparing white and Asian candidates only. The data on 13 843 candidates (10 974 white and 2675 Asian) came from 10 reports, comprising 16 datasets,^{27 33 35 42 44 46} and included raw data from Lumb and Vail³² and three papers by Woolf et al.^{20 37 39} The results were similar ($d = -0.40$, -0.51 to -0.28 ; $I^2 = 80\%$).

None of the funnel plots for these meta-analyses showed any sign of publication bias.

Summary of adjusted effects

Thirteen of the 23 reports gave details of figures adjusted for various other factors (table 5). Eleven of the 13 showed unadjusted significant effects of ethnicity on outcomes. Of the 11, only Ricketts et al found that adjusting for covariates removed a previously significant effect of ethnicity on outcomes.⁴⁶ The covariates in that study were sex, disability, year (their outcome measure was a progress multiple choice test taken by students across all five years), and three interaction terms. The 10 other reports that showed simple effects of ethnicity also found significant effects of ethnicity on outcomes after adjustment for sex. Kilminster et al reported no significant interaction between the effects of sex and ethnicity.⁴¹

The following studies all found a significant effect of ethnicity on outcomes: five studies that adjusted for

previous exam performance,^{32 34-36 40} one study in UK candidates speaking English as a first language,³³ two studies that adjusted for school type,^{32 40} and one study that adjusted for socioeconomic group.³²

DISCUSSION

Principal findings

This meta-analysis shows that doctors and medical students of non-white ethnicity underperform academically compared with their white counterparts. The effect was significant ($P < 0.001$) and, using the terminology of Cohen⁵⁰, was of medium magnitude ($d = -0.42$). It should be remembered in interpreting such effects that Cohen's d describes differences in terms of the means, whereas proportions in the tails of a distribution can be much larger. To give an example, on a typical exam with a mean score in white candidates of 60 (SD 5), an effect size of $d = -0.42$ would mean that non-white candidates would score an average of 57.9. If the pass mark were 50, then, under certain statistical assumptions, 2.3% of white and 5.6% of non-white candidates would fail, making the odds of failure in non-white candidates 2.5 times higher than for white candidates. For those studies reporting pass/fail outcomes, we found an overall odds ratio of 2.92 ($P < 0.001$).

Separate meta-analyses of undergraduate assessments ($d = -0.42$; $P < 0.001$), postgraduate assessments ($d = -0.38$; $P < 0.001$), machine marked written assessments ($d = -0.35$; $P < 0.001$), practical clinical assessments ($d = -0.42$; $P < 0.001$), assessments with pass/fail outcomes ($d = -0.59$; $P < 0.001$), and assessments with continuous outcomes ($d = -0.38$; $P < 0.001$) all showed similar effects in the same direction. The comparison of white and Asian candidates only also showed a similar result ($d = -0.40$; $P < 0.001$). Though there were varying amounts of heterogeneity in the meta-analyses, it is clear that the finding of an ethnic difference in assessment outcomes is both consistent and persistent. Of 36 datasets included in the meta-analysis of all studies, 35 showed a negative effect of non-white ethnicity and in 25 the effect was significant. None showed a significant positive effect of non-white ethnicity; one showed no difference between the white and non-white groups.

Strengths and weaknesses

Our meta-analysis contained data from nearly 24 000 candidates. It provides evidence that ethnic differences in postgraduate attainment exist independently of the known lower performance in postgraduate examinations of overseas candidates.^{9 51 52} The separate analysis of machine marked written assessments and practical assessments allowed us to investigate possible effects of examiner bias and verbal communication skills on ethnic differences in attainment. That an ethnic attainment gap was found in both machine marked and face to face assessments suggests that those factors are unlikely to be primarily responsible, although effects might still be present. Our summary of adjusted results enables us to begin to look at some of the possible confounders and showed that differences in attainment are unlikely to be more prominent in men or women. The

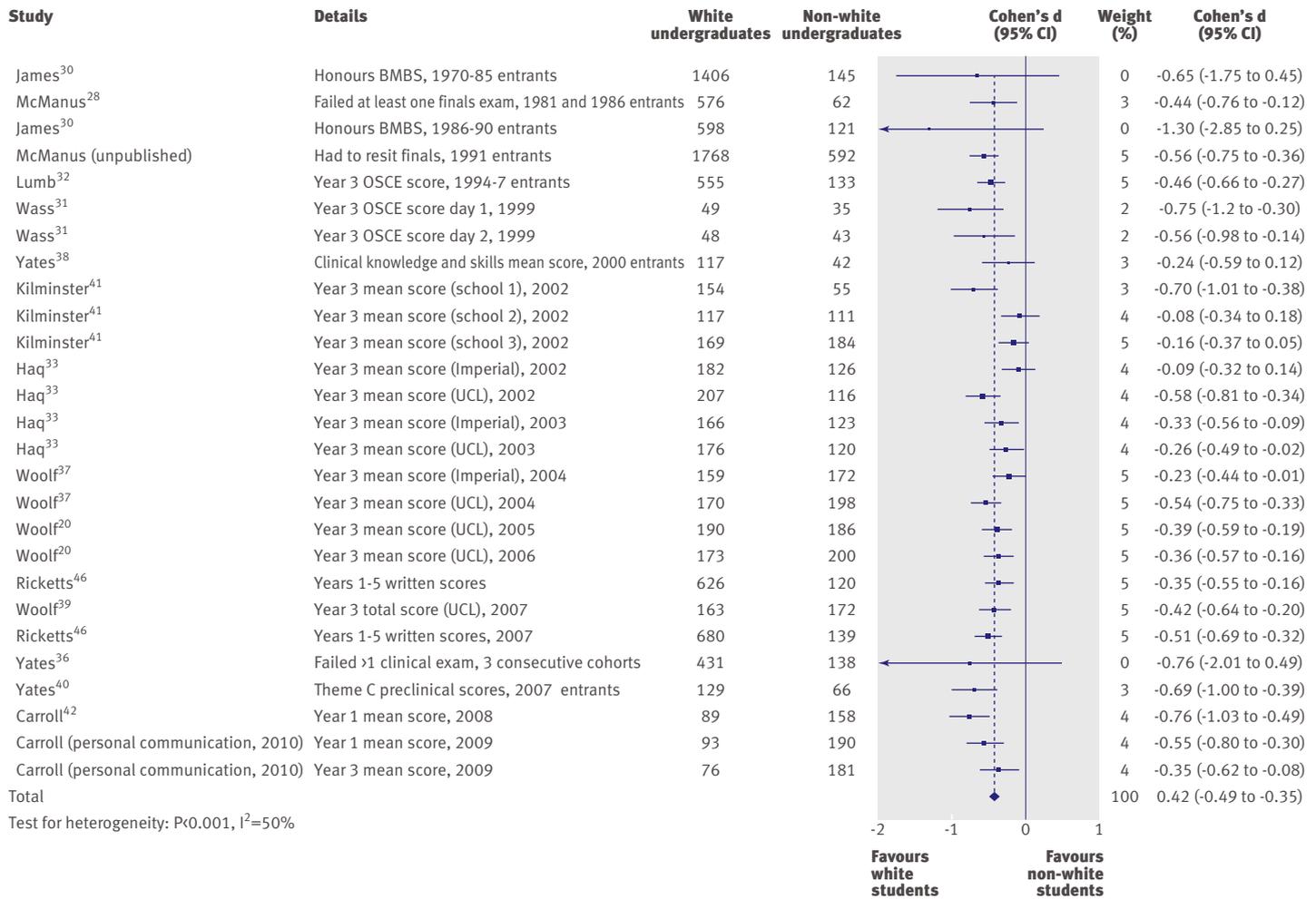


Fig 4 | Results of meta-analysis of 16 reports (27 datasets) on non-white ethnicity on performance in undergraduate medical students, showing significantly poorer performance in non-white students

fact that our study includes data from postgraduate as well as undergraduate assessments highlights that this effect is not restricted to medical students but affects practising doctors.

Funnel plots showed no apparent publication bias. As we have a particular interest in this issue we are often asked to analyse unpublished assessment data to check for ethnic differences, and, when possible, we included those data to reduce publication bias. This approach can introduce a different type of selection bias, and so we suggest that all UK medical schools and Royal Colleges should analyse their assessment data for ethnic differences and publish the results.

Interpretation of the results is somewhat tempered by heterogeneity, particularly in the postgraduate sample. The meta-analysis of undergraduate results had an I^2 of 50%, which can be considered as “moderate” and, according to Higgins and Thompson’s tentative estimations, is unlikely to be of much concern.⁵³ The postgraduate data, however, had a high heterogeneity of $I^2=89%$, and there is clearly unexplained variation in the effect size here. Forest plots did not suggest any simple explanations, and, while we were unable to conduct a meta-analysis of effects adjusted for confounders

(because of a lack of reported data and variation in the confounders adjusted for), our summary of adjusted effects in table 4 does not point to a single explanatory variable. The heterogeneity in the postgraduate studies could have been because of varying assessment formats or reliabilities. For example, the shortlisting scores reported in Brown et al⁴⁵ are different in many ways from the results of a tightly controlled machine marked MRCGP exam.⁴⁴ It might also have reflected differences in the comparison groups between studies. One of the largest studies in the meta-analyses compared Asians with non-Asians.²⁷ While the largest proportion of the non-Asian group was probably white, that group would also have included non-white candidates; removal of this study from the analyses, however, did not significantly alter the findings.

There are other possible reasons for heterogeneity in both undergraduate and postgraduate samples. The proportion of white and non-white students was different in the older studies compared with the newer ones. To explore whether this affected the heterogeneity of the results, we conducted a post hoc meta-analysis of 14 studies (23 datasets) from candidates who can reasonably be expected to have entered medical school in

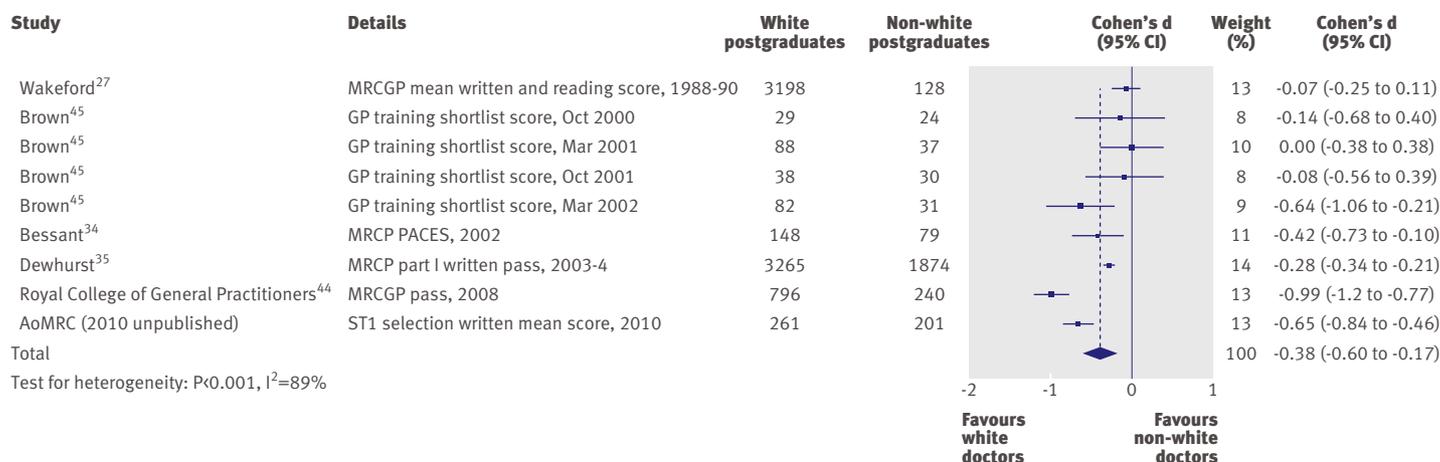


Fig 5 | Results of meta-analysis of 16 reports (27 datasets) on non-white ethnicity on performance in postgraduate medical studies, showing significantly poorer performance in non-white candidates

or after 1997, when the number of medical students in the UK expanded rapidly.⁵⁴ It showed similar results ($d = -0.44, -0.55$ to -0.35 ; $P < 0.001$; $I^2 = 72\%$). This also shows that the ethnic gap is a feature of current medical education.

The proportions of candidates from different minority ethnic groups might also have varied between reports, and it is possible that differential performance between minority ethnic groups (as is found in school children, where, for example, Indians achieve higher grades than Bangladeshis^{55,56}) might have resulted in heterogeneity. It is also important to consider how the variation in methods for obtaining ethnic data (such as self report, use of name, and photograph; in three studies the method was unknown), as well as the exclusion of participants whose ethnicity was unknown, could have affected the reliability of the results.

Our meta-analysis comparing white and Asian candidates was not subtle enough to distinguish between different groups of Asians and showed similar results to the overall meta-analysis, including an I^2 of 80%. To tease out these differences would need studies with larger sample sizes. At present we know of only one UK study within medicine with a large enough sample size to look for differences between several minority ethnic groups, and it found no significant differences between them in terms of pass rates³⁵; that study, however, was not able to distinguish between, say, Bangladeshis and Pakistanis, let alone any finer categorisations.

Another weakness of our study is that most of the undergraduate reports in this meta-analysis came from London and Nottingham medical schools (though the postgraduate studies probably included graduates of all UK medical schools). Medical schools vary in their curriculums, teaching methods, and proportions of minority ethnic students from various groups and their graduates vary in postgraduate attainment.⁵⁷ So, though the effect was clear in the five non-London/non-Nottingham schools in the study, care should be taken when generalising our

findings outside England. Similarly, the postgraduate reports were mostly of general practitioners and physicians, and there were no studies of surgeons, for example. We could not therefore study possible variation in ethnic differences across specialties and grades.

While we were able to examine in a limited way the impact of various covariates (particularly sex) on the ethnic gap in attainment, we were not able to do a formal analysis of the relation between ethnicity, socioeconomic status, and attainment. While most medical students in the UK are from the highest socioeconomic groups,^{6,7} this might well vary as a function of ethnicity and medical school and therefore needs exploring.

Strengths and weaknesses in relation to other studies

To our knowledge, the only other systematic review of the evidence at undergraduate level was carried out by Ferguson and colleagues.² Those authors also found that non-white ethnicity negatively predicted undergraduate performance, but their review contained only one UK study, the others coming mainly from the United States. In postgraduate terms, we do not know of any systematic reviews of qualified doctors' academic performance in relation to ethnicity.

Most reports retrieved in our search that did not meet the inclusion criteria also found that undergraduates from ethnic minorities underperformed.^{1,2,5,47,48} In contrast, Arulampalam and colleagues found that Indian women and non-white non-Indian ("other") students were less likely than white students to drop out of medicine in the first year.²⁶ Flint and colleagues reported no significant ethnic differences in terms of GP placement success for UK graduates in 2004 and 2008, although white UK candidates did achieve higher scores in knowledge tests in 2009.⁴³

The question of attainment before medicine is also important. In a national sample, McManus et al found that non-white students tended to enter medical school with slightly lower grades in school leaving exams (effect size $d = -0.10$).⁵⁸ There is also evidence that applicants⁵⁹⁻⁶² and entrants⁴⁰ from ethnic minorities

Table 5 Summary of main ethnicity effects and ethnicity effects adjusted for covariates in UK medical students

Author	Main outcome measure	Unadjusted effect of ethnicity significant	Effect of ethnicity significant after adjustments for covariates
James ³⁰	Passed BMBS with honours	1970-85: no; 1986-90: no	1970-85 cohort: no effect after adjustment for O level chemistry and total predicted score for all A levels (1970-85 cohort). 1986-90 cohort: adjustments not reported
Yates ³⁸	Combined clinical knowledge and skills	No	No effect after adjustment for fee status, sex, entry qualifications, previous medical school performance
Ricketts ⁴⁶	MCQ score	Yes (both cohorts)	No effect after adjustment for main effects of disability, sex, and year and interaction effects of year×disability, year×sex, year×ethnicity, disability×sex, disability×ethnicity, and sex×ethnicity, though no interactions were significant (2006 and 2007)
McManus ²⁸	Failed at least one finals exam	Yes	Effect after adjustment for sex, and analysis restricted to UK nationals only
Lumb ³²	OSCE score	Yes	Effect after adjustment for GCSE points, A level points, subjects studied at A level, application form assessment scores in 4 subsections, sex, socioeconomic group, school type, applicant category (school leaver, 2nd application for medicine, deferred entry, mature entrant), whether interviewed
Bessant ³⁴	Passing PACES	Yes	Effect after adjustment for sex, age, having BSc, having been on previous PACES course, having passed MRCP(UK) part 2 first time, work experience
Dewhurst ³⁵	MRCP(UK) pass rate	Yes	Effect after adjustment for sex and attempt number in each of part 1, part 2, PACES
Yates ³⁶	Failed at least one clinical exam	Yes	Effect after adjustment for sex, A level results, late offer
Yates ⁴⁰	Preclinical scores	Yes, theme C only	Effect in theme C, after adjustment for sex, UKCAT score, and school type
AoMRC (unpublished, 2010)	Written clinical problem solving test	Yes	Effect after adjustment for sex, age, country of primary medical qualification, and preferred speciality
Woolf ³⁷	Combined OSCE and MCQ scores	Yes (both cohorts)	Effect after adjustment for sex (both cohorts)
Woolf ³⁹	Combined OSCE and MCQ scores	Yes	Effect after adjustment for sex
Woolf ²⁰	End of year scores	Yes (both cohorts)	Effect after adjustment for sex (both cohorts)
Kilminster ⁴¹	OSCE and MCQ scores	School 1: yes; school 2: no; school 3: no	Adjustments not reported, though sex and ethnicity interactions were not significant
Haq ³³	MCQ and OSCE scores	Yes	Adjustments not reported, but analysis restricted to Asian and white UK candidates with English as first language
Wakeford ²⁷	MRCGP pass rates	Yes	Adjustments not reported
Wass ³¹	OSCE score	Yes	Adjustments not reported
Brown ⁴⁵	Shortlisting scores	Oct 2000: no; Mar 2001: no; Oct 2001: no; Mar 2002: yes	Adjustments not reported
Brown ⁴⁹	Shortlisting outcome	Yes	Adjustments not reported
RCGP ⁴⁴	Clinical skills assessment and applied knowledge test pass rates	Yes	Adjustments not reported
Carroll ⁴²	Written score	Yes	Adjustments not reported
Carroll (personal communication, 2010)	Combined written and clinical scores	Yes	Adjustments not reported
McManus (unpublished)	Passed/failed finals (fail, retake year, retake one exam, pass, merit or distinction)	Yes	Adjustments not reported

MCQ=multiple choice questions; PACES=practical assessment of clinical examination skills; OSCE=objective structured clinical examination.

have lower results on the UKCAT (aptitude test used to select medical students) than white students. Unlike A levels, however, which are known to be reasonably good predictors of medical school performance,⁶³ the evidence so far suggests that UKCAT is not good predictor.^{40 64} Our study was concerned with performance at medical school and beyond and therefore did not consider attainment before medicine. We did report the results of three studies from two medical schools that examined whether adjusting for previous grades removed the effect of ethnicity on attainment at medical school, and they found it did not.^{32 40 46} Once again, larger scale longitudinal studies would be required to examine this important issue further. The question of why medical students from ethnic minorities have, on average, slightly lower entry

qualifications than white students is a related question that also requires attention.

Unanswered questions and further research

Ethnic differences in attainment seem to be a consistent feature of medical education in the UK, being present across medical schools, exam types, and undergraduate and postgraduate assessments, and have persisted for at least the past three decades. They cannot be dismissed as atypical or local problems. This is an uncomfortable finding, with good reason. While exam performance is by no means the only marker of good performance as a doctor or medical student, the fact remains that without passing finals, medical students cannot become doctors, and without passing postgraduate exams, it is much harder for doctors to progress in

WHAT IS ALREADY KNOWN ON THIS TOPIC

A third of all UK medical students and junior doctors are from minority ethnic groups

A 2002 review of the factors influencing medical school success found evidence of underperformance in minority ethnic candidates, but included just one report of UK data

UK universities and the NHS are legally required to monitor the admission and progress of students and the recruitment and career progression of staff by ethnic group

WHAT THIS STUDY ADDS

UK trained doctors and medical students from minority ethnic groups tend to underperform academically compared with their white counterparts

Ethnic differences are unlikely to be primarily caused by examiner bias or candidate communication skills because similar effects are found in machine and examiner marked assessments

a medical career. That exam results vary by ethnicity is therefore extremely important and requires attention. Although ethnicity is clearly related to exam performance, what is not clear is why that might be.^{20,33,37} This meta-analysis allows us to move on from publishing the effects of ethnicity on exams in single medical schools or individual college membership exams to exploring the reasons for this gap in attainment and what might be done about it.

More detailed information is needed to track the attainment gap, and further research is needed into its causes, which, like ethnic differences in achievement in primary and secondary education, are probably complex and multifactorial.⁶⁵ To begin to address the problem, it needs to be recognised as a shared problem. A proper approach will be for all medical schools and Royal Colleges to analyse their assessment results by ethnic group and place their results in the public domain and to encourage educational researchers to examine possible mechanisms, such as stereotype threat,⁶⁶ and test interventions for improvements.³⁹ Medical students and doctors from all ethnic groups will need to be involved in this process. Without these actions, it will be a struggle to ensure a fair and just method of training and assessing our future and current doctors.

We thank Mark Carroll, David James, Sue Kilminster, Trudie Roberts, and Andy Lumb for sharing their raw and unpublished data with us.

Contributors: ICM conceived the idea for the study, which was further developed with KW and HWWP. KW and HWWP decided on inclusion and exclusion criteria. KW and ICM searched for reports and extracted data. All authors double extracted data and contributed to the statistical analysis and interpretation of results. KW prepared the first draft of the paper, and all authors contributed to subsequent drafts and agreed on the final report. All contributors had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. KW is guarantor.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the

previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: The study was covered by the exemptions of the UCL research ethics committee <http://ethics.grad.ucl.ac.uk/exemptions.php>. Consent from participants was not obtained but the presented data are anonymised and risk of identification is low.

Data sharing: Datasets are available from the corresponding author at k.woolf@medsch.ucl.ac.uk.

- Dillner L. Manchester tackles failure rate of Asian students. *BMJ* 1995;310:209.
- Ferguson E, James D, Madeley L. Factors associated with success in medical school: systematic review of the literature. *BMJ* 2002;324:952-7.
- Broecke S, Nicholls T. Ethnicity and degree attainment. Research report RW92. Department for Education and Skills, 2007.
- Fielding A, Charlton C, Kounali D, Leckie G. Degree attainment, ethnicity and gender: interactions and the modification of effects. A quantitative analysis. Equality Challenge Unit, Higher Education Academy, 2008.
- Richardson JTE. Degree attainment, ethnicity and gender: a literature review. Higher Education Academy, 2008.
- McManus IC. Measuring participation in UK medical schools: social class data are problematic to interpret. *BMJ* 2004;329:800-1.
- Seyan K, Greenhalgh T, Dorling D. The standardised admission ratio for measuring widening participation in medical schools: analysis of UK medical school admissions by ethnicity, socioeconomic status, and sex. *BMJ* 2004;328:1545-6.
- Department of Health. Sharing the challenge, sharing the benefits: equality and diversity in the medical workforce. DH Publications, 2004.
- British Medical Association. Examining equality: a survey of royal college examinations. BMA, 2006.
- Roberts C, Sarangi S, Southgate L, Wakeford R, Wass V. Oral examinations—equal opportunities, ethnicity, and fairness in the MRCP. *BMJ* 2000;320:370-4.
- UCAS Statistics. Annual UK dataset for ethnicity and social class. 2009. www.ucas.ac.uk/about_us/stat_services/stats_online/annual_datasets_to_download/.
- NHS Staff Statistics. Medical and dental staff 1999-2009. NHS Information Centre, 2011. www.ic.nhs.uk/statistics-and-data-collections/workforce/nhs-staff-numbers/nhs-staff-1999-2009-medical-and-dental.
- Goldacre MJ, Davidson JM, Lambert TW. Country of training and ethnic origin of UK doctors: database and survey studies. *BMJ* 2004;329:597.
- Race Relations (Amendment) Act 2000 (c 34). Stationery Office, 2000.
- Higher Education Funding Council for England. HEFCE equality scheme. 2007. www.hefce.ac.uk/pubs/hefce/2007/07_01/07_01.doc.
- Gill PS. General practitioners, ethnic diversity and racism. In: Coker N, ed. Racism in medicine: an agenda for change. King's Fund, 2001:99-120.
- Senior PA, Bhopal R. Ethnicity as a variable in epidemiological research. *BMJ* 1994;309:327-30.
- Malik K. Strange fruit: why both sides are wrong in the race debate. One World Publications: Oxford, 2008.
- Platt L. Ethnicity and child poverty. Research report No. 576. Department for Work and Pensions, 2009. <http://research.dwp.gov.uk/asd/asd5/rports2009-2010/rrep576.pdf>
- Woolf K. The academic underperformance of medical students from ethnic minorities. [PhD thesis]. University of London, 2009.
- Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ* 2005;331:1064-5.
- Hedges LV. Distribution theory of Glass's estimator of effect size and related estimators. *J Educ Stat* 1981;6:107-28.
- Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med* 2000;19:3127-31
- McManus IC, Richards P. Prospective survey of performance of medical students during preclinical years. *BMJ* 1986;293:124-7.
- Yates J, James D. Predicting the "strugglers": a case-control study of students at Nottingham University Medical School. *BMJ* 2006;332:1009-13.
- Arulampalam W, Naylor RA, Smith JP. Dropping out of medical school in the UK: explaining the changes over ten years. *Med Educ* 2007;41:385-94
- Wakeford R, Farooqi A, Rashid A, Southgate L. Does the MRCPG examination discriminate against Asian doctors? *BMJ* 1992;305:92-4.

- 28 McManus IC, Richards P, Winder BC, Sproston KA. Final examination performance of medical students from ethnic minorities. *Med Educ* 1996;30:195-200.
- 29 McManus IC, Richards P, Winder BC, Sproston KA, Styles V. Medical school applicants from ethnic minorities: identifying if and when they are disadvantaged. *BMJ* 1995;310:496-500.
- 30 James D, Chilvers C. Academic and non-academic predictors of success on the Nottingham undergraduate medical course 1970-1995. *Med Educ* 2001;35:1056-64.
- 31 Wass V, Roberts C, Hoogenboom R, Jones R, Van der Vleuten C. Effect of ethnicity on performance in a final objective structured clinical examination: qualitative and quantitative study. *BMJ* 2003;326:800-3.
- 32 Lumb AB, Vail A. Comparison of academic, application form and social factors in predicting early performance on the medical course. *Med Educ* 2004;38:1002-5.
- 33 Haq I, Higham J, Morris R, Dacre J. Effect of ethnicity and gender on performance in undergraduate medical examinations. *Med Educ* 2005;39:1126-8.
- 34 Bessant R, Bessant D, Chesser A, Coakley G. Analysis of predictors of success in the MRCP (UK) PACES examination in candidates attending a revision course. *Postgrad Med J* 2006;82:145-9.
- 35 Dewhurst N, McManus IC, Mollon J, Dacre J, Vale JA. Performance in the MRCP(UK) Examination 2003-4: analysis of pass rates of UK graduates in relation to self-declared ethnicity and gender. *BMC Med* 2007;5:8.
- 36 Yates J, James D. Risk factors for poor performance on the undergraduate medical course: cohort study at Nottingham University. *Med Educ* 2007;41:65-73.
- 37 Woolf K, Haq I, McManus IC, Higham J, Dacre J. Exploring the underperformance of male and minority ethnic medical students in first year clinical examinations. *Adv Health Sci Educ Theory Pract* 2008;13:607-16.
- 38 Yates J, Smith J, James D, Ferguson E. Should applicants to Nottingham University Medical School study a non-science A level? A cohort study. *BMC Med* 2009;9:5.
- 39 Woolf K, McManus IC, Gill D, Dacre J. The effect of a brief social intervention on the examination results of UK medical students: a cluster randomised controlled trial. *BMC Med* 2009;9:35.
- 40 Yates J, James D. The value of the UK Clinical Aptitude Test in predicting pre-clinical performance: a prospective cohort study at Nottingham Medical School. *BMC Med* 2010;10:55.
- 41 Kilminster S, Boursicot K, Roberts T. Do gender and ethnicity affect medical students' examination results? *Med Teach* in press.
- 42 Carroll M, Mackenzie J. Why do "non-white" students under-achieve academically? Association for the Study of Medical Education (ASME) Annual Scientific Meeting, Cambridge, July 2010.
- 43 Plint S, Patterson F, Jackson N, Gregory S. Equal opportunities in selection: an evaluation of UK national recruitment for general practice. Association for Medical Education Europe Annual Conference, Malaga, 2009.
- 44 nMRCGP: report of results 2008. Royal College of General Practitioners, 29 September 2008. www.educatingthefuturegp.co.uk/useful%20files/nMRCGP%20interim%20report.pdf.
- 45 Brown CA, Wakefield SE, Bullock AD. The selection of GP trainees in the West Midlands: second audit of assessment centre scores by ethnicity and country of qualification. *Med Teach* 2003;25:649-53.
- 46 Ricketts C, Brice J, Coombes L. Are multiple choice tests fair to medical students with specific learning difficulties? *Adv Health Sci Educ* 2010;15:265-75.
- 47 Calvert MJ, Ross NM, Freemantle N, Xu Y, Zvauya R, Parle JV. Examination performance of graduate entry medical students compared with mainstream students. *J R Soc Med* 2009;102:425-30.
- 48 Kitchen RE, Sahota RS, Stanley AG, Crick H. Effects of demographic factors on student performance in basic science exams. [Poster presentation.] Association for the Study of Medical Education. Annual Scientific Meeting: Leicester, 2008. www.asme.org.uk/images/pdfs/asm_posters.pdf.
- 49 Brown CA, Wakefield SE, Bullock AD. The selection of GP trainees in the West Midlands: audit of assessment centre scores by ethnicity and country of qualification. *Med Teach* 2001;6:605-9.
- 50 Cohen J. Statistical power analysis. *Curr Dir Psychol Sci* 1992;1:98-101.
- 51 Tyrer SP, Leung W-C, Smalls J, Katona C. The relationship between medical school of training, age, gender and success in the MRCPsych examinations. *Psychiatr Bull* 2002;26:257-63.
- 52 Oyeboode F, Furlong E. MRCPsych examinations: cumulative results 1997-2002. *Psychiatr Bull* 2007;31:61-4.
- 53 Higgins JPT, Thompson SG, Deeks JJ. Measuring inconsistency in meta-analyses *BMJ* 2003;327:557-60.
- 54 Elston MA. Women in medicine: the future. Royal College of Physicians, 2009.
- 55 Ethnicity and education: the evidence on minority ethnic pupils. RTP01-05. Department for Education and Skills, 2005. www.education.gov.uk/publications//eOrderingDownload/DFES-0208-2006.pdf.
- 56 Bhattacharyya G, Ison L, Blair, M. Minority ethnic attainment and participation in education and training: the evidence. RTP01-03. Department for Education and Skills, 2003. www.teachernet.gov.uk/_doc/6137/RTP01-03.pdf.
- 57 McManus IC, Elder AT, de Champlain A, Dacre JE, Mollon J, Chis L. Graduates of different UK medical schools show substantial differences in performance on MRCP(UK) Part 1, Part 2 and PACES examinations. *BMC Med* 2008;6:5.
- 58 McManus IC, Woolf K, Dacre JE. The educational background and qualifications of UK medical students from ethnic minorities. *BMC Med* 2008;8:21.
- 59 UK Clinical Aptitude Test. UKCAT annual report 2008. UKCAT, 2008. www.ukcat.ac.uk/pdf/Annual%20Report%202008.pdf.
- 60 UK Clinical Aptitude Test. UKCAT annual report 2007. UKCAT, 2007. www.ukcat.ac.uk/pdf/UKCAT%20Annual%20Report%202007.pdf.
- 61 UK Clinical Aptitude Test. UKCAT annual report 2006. UKCAT, 2006. <http://www.ukcat.ac.uk/pdf/UKCAT%20Annual%20Report%202006%20v2.pdf>.
- 62 James D, Yates J, Nicholson S. Comparison of A level and UKCAT performance in students applying to UK medical and dental schools in 2006: cohort study. *BMJ* 2010;340:c478.
- 63 McManus IC, Powis DA, Wakeford R, Ferguson E, James D, Richards P. Intellectual aptitude tests and A levels for selecting UK school leaver entrants for medical school. *BMJ* 2005;331:555-9.
- 64 Lynch B, MacKenzie R, Dowell J, Cleland J, Prescott G. Does the UKCAT predict year 1 performance in medical school? *Med Educ* 2009;43:1203-9.
- 65 Report of the committee of enquiry into the education of children from ethnic minority groups: education for all (Swann report). HMSO, 1985. www.educationengland.org.uk/documents/swann/swann03.html.
- 66 Steele CM, Aronson J. Stereotype threat and the intellectual test performance of African Americans. *J Pers Soc Psychol* 1995;69:797-811.

Accepted: 29 December 2010