Independent Review of the Membership of the Royal College of General Practitioners (MRCGP) examination

Aneez Esmail
Professor of General Practice

Chris Roberts
Professor of Biostatistics

The University of Manchester

September 2013

(18th September 2013)

This review was commissioned by the GMC in order to identify if there was any evidence of discrimination against International Medical Graduates (IMG) or British Black and Minority Ethnic graduates (BME) in the MRCGP examination that is set by the Royal College of General Practitioners (RCGP). The terms of reference of the review are attached as Appendix 1.

Context and background

It has been known for some time that there is a difference in pass rates for IMG and BME candidates sitting the MRCGP examination[1] from UK graduates and white candidates. These differences in pass rates are not restricted to examinations set by the RCGP. The Royal College of Psychiatrists[2] and the Royal College of Physicians[3] have commissioned and published data that has highlighted the increased failure rate of IMG and British BME doctors in their postgraduate examinations. Differences in pass rates between indigenous and international medical graduates have also been highlighted in postgraduate examinations in Australia and the USA[4;5].

Explaining and understanding the differential pass rates between IMG and British graduates in the MRCGP is not simply an academic exercise. The NHS continues to depend on IMG to provide a significant contribution to the workforce in order to meet its staffing needs. This dependence increased as a result of the NHS Plan 2000, which resulted in a significant expansion of NHS services and the concomitant increase required in the medical workforce that could not be met solely by the recruitment of British trained graduates. Between 2002 and 2012 there has been a 25% increase in the headcount of general practitioners. The proportion of UK graduates in the GP workforce was 77% in 2012, suggesting that non-UK qualifiers now account for almost a quarter of the GP workforce. In 2012 there were estimated to be nearly 10,217 non-UK graduates working as GPs[6]. How this group enters general practice and the potential barriers they face is therefore hugely important to the future of the GP workforce. The situation is likely to change in the future due to UK government migration policies which has had the effect of restricting the overall number of IMG.

The increase in the number of non-UK qualifiers in General Practice has taken place against a background of significant regulatory changes governing the entry and qualifying requirements for the different medical specialties. These changes were developed as part of the Modernising Medical Careers (MMC) initiative started by the Department of Health in 2003. The initiatives associated with MMC have undergone many changes since the programme was first introduced in 2007 but certain key features remain.

The GMC regulates the standards and assessments required for completion of specialty training and inclusion in the GMC GP Register for those following a training pathway requires a Certificate of Completion of Training (CCT)[i]

[i] Or a certificate of Eligibility for General Practice(CEGPR) via the combined programme route

.

GP training

Entry into specialty training for General Practice is now organised centrally through the National Recruitment Office for General Practice Training (NRO). Training usually takes 3 years (although 2/3 of the Scottish GP programmes are 4 years in duration) after the completion of the two years of Foundation Year training. In 30% of Foundation programmes in Scotland and 40% in England the training will include 4 months in a GP environment. The three year GP specialty training will usually consist of 18 months working in a general practice under the supervision of a GP trainer, and the remaining time in educationally approved hospital posts relevant to the work of a GP, such as paediatrics, obstetrics, gynaecology, psychiatry, medicine or accident and emergency. During the time in general practice the GP registrar will follow the approved curriculum - learning how general practice is organised and managed and will see patients both in the surgery and in their homes.

Pathways for International Medical Graduates (IMG) entering General Practice are likely to be different. Although some IMG complete Foundation training posts, the majority will come through the full registration route. In order to be considered for full registration, IMG need to have completed an English Language Capability Test (IELT) by scoring a minimum of 7[ii] in all components (speaking, reading, writing and listening) of the academic version IELTS. They also have to sit the Professional and Linguistic Assessment Board (PLAB) examination. This consists of two parts and Part 2 is an Objective Structured Clinical Examination, similar in some aspects to the Clinical Skills Assessment examination of the MRCGP (see below). Once they have achieved foundation competences either via F2 year or of demonstrating equivalence, IMG would be eligible to apply for GP specialist training through the NRO.

At the end of training, satisfactory completion of the MRCGP examination is a pre-requisite for the award of a CCT. The concerns regarding the failure rate of IMG and BME British graduates are therefore also important since failure to pass the MRCGP examination effectively means that a person cannot work in general practice in any capacity. The number of attempts at each component of the MRCGP is now restricted to four attempts, which is within the standard which has been agreed with the Academy of Medical Royal Colleges and set by the GMC. It is therefore a high stakes examination, with failure ultimately restricting the ability to work in general practice. From the perspective of the GMC and the Medical Colleges, it is important for patient safety that a standard is met and this may mean failure in a summative examination.

The MRCGP examination

The MRCGP comprises three separate components: an Applied Knowledge Test (AKT), a Clinical Skills Assessment (CSA) and Workplace Based Assessment (WPBA), each of which tests different competences using validated assessment methods and which together cover the spectrum of knowledge, skills, behaviours and attitudes defined by the GP Specialty Training curriculum. This version of the examination was introduced in 2007 and approved by the regulator for postgraduate examinations at that time – the Postgraduate Medical Education Training Board (PMETB)

[ii] The score requirements changed in October 2010; prior to then the requirement was a score of 7 overall and a 7 in speaking and at least a 6 in the other 3 domains.

The Applied Knowledge Test is a summative assessment of the knowledge base that underpins independent general practice in the United Kingdom. It is a machine marked multiple choice examination. Candidates currently pay a fee of £454[iii] to sit this examination and are allowed a maximum of four attempts to sit the examination. It can be sat during or after the second year of GP training.

The Clinical Skills Assessment (CSA) is a summative assessment of a doctor's ability to integrate and apply clinical, professional, communication and practical skills appropriate for general practice. The format of the examination simulates a typical NHS surgery clinic and assesses a range of scenarios from general practice. Candidates currently pay £1525[iv] to sit this examination and are allowed a maximum of four attempts to sit their examination. It can be sat during or after the third year of GP training.

The Workplace Based Assessments (WPBA) defined by the curriculum evaluate the trainee's progress in areas of professional practice best tested in the workplace and is a continuous and formative assessment carried out by the designated GP trainer (who is the educational supervisor for the complete GP programme). Supervision and formative assessments are also completed by clinical supervisors in hospital posts. It is overseen by the postgraduate deanery[v] in which the training is taking place and to ensure national consistency, the RCGP quality assures the WPBAs through sampling of ARCPs. Completion of satisfactory workplace based assessments is also a requirement for obtaining a Certificate of Completion of Training (CCT).

A qualitative picture of a trainee's performance in training is built up using workplace based assessments, educational and clinical supervisors' reports based on observation and examination performance and is reviewed annually through a process called the Annual Review of Competence Progression (ARCP). It is possible to receive an unsatisfactory ARCP if it is deemed that the trainee shows an insufficient and sustained lack of progress. Trainees may sometimes require additional time in training or simply targeted training in specific areas. The GMC has commissioned a separate study looking at the outcome of ARCPs and preliminary evidence suggests that unsatisfactory outcomes are more common for IMG than UK graduates across most specialties.

Changes in the method of marking the CSA in 2010

The MRCGP underwent significant changes in 2010, specifically in relation to the method of marking the CSA component. It is relevant to understand these changes because it has determined the time frame for the review and why we have not considered data from 2007-2010.

Prior to 2010, during the CSA, the candidate was assessed undertaking 13 clinical scenarios (cases); although all 13 cases were marked only 12 cases were counted towards the candidate's overall score. The 13th case was used to pilot new cases and did not contribute to the candidate's overall mark. When it was introduced in 2007, the passing standard for the CSA had been based on a

[iii] The fee is higher (£506) for those that are not RCGP Associates in training (AiT)
[iv] The fee is higher (£1694) for those that are not AiTs
[v] In April 2013 Postgraduate Deaneries became Local Education and Training Boards (LETBs) in England. This report was commissioned prior to this change and therefore only makes reference to LETBs in the recommendations.

'number of cases to pass' (N2P) methodology. The Panel of Examiners of the RCGP CSA Core Group felt that it had been impossible to fine tune this to make allowances for daily variability in the difficulty of the mix of cases, compensation between performance on different cases or the effects on the passing standard of increasing familiarity of the examination by both trainees and the examiners.

The RCGP therefore felt that there needed to be a change to the standard setting methodology to increase confidence levels to ensure that the RCGP was passing doctors who were competent and safe. The RCGP felt that the reliability and fairness of the CSA examination could be improved by introducing a standard setting method that took account of the pass/fail borderline thereby improving the reliability of the assessment and compensating between cases and domains in setting the standard. The RCGP had already been asked to review their method of standard setting in 2008. There was also a view by examiners, GP trainers and patient representatives on the RCGP examination board, that a passing standard of eight out of 12 cases was too low. Put simply, under N2P methodology, eight marginal passes with four clear fails was a pass whilst seven clear passes with five marginal fails was an overall fail.

Under the borderline group methodology the examiners, as well as marking against domains (one examiner marks each candidate they observe on a case giving it one of four grades and each candidate is also graded against three domains - Data Gathering, Clinical Management and Interpersonal Skills), makes a further standard setting judgment, rating the candidate as pass, borderline or fail. For each case the marks of those candidates marked as borderline are averaged. These averaged borderline scores are then aggregated across all 13 cases to create the "cut score", i.e. the approximation between a passing and a failing score. The final, actual pass mark has an adjustment to the overall cut score to take account of the measurement error inherent in any assessment process of this kind. It is the application of this adjustment, known as the standard error of the mean (SEM), which is controversial. The GMC have approved that the SEM methodology of standard setting with a narrow range fulfils its standards for curricula and assessment systems.

The RCGP therefore introduced a borderline group marking methodology and included the 13th clinical scenario as a marked case. There was extensive consultation with international experts, pilot testing and statistical modelling to assess the impact of these changes. The borderline method of marking examinations like the CSA is widely used both internationally and in the UK. It is the standard method in some medical schools when assessing students in clinical examinations (widely known as Objective Structured Clinical Examinations). It is also used along with other marking schemes by the GMC in marking the PLAB Part 2 examination. These changes were approved by the GMC in 2010.

The role of the GMC

The General Medical Council (the GMC) is the independent regulator for doctors in the UK and the competent authority for awarding qualifications to those who satisfactorily complete training in one

of the approved specialties which includes General practice. Their primary purpose is to protect, promote and maintain the health and safety of the public by ensuring proper standards in the practice of medicine. It is the GMC who commissioned this review as part of their responsibility to ensure that they were fulfilling their responsibilities for educational standards.

The GMC's Education strategy 2011-2013 sets out a series of key aims which includes setting and assuring standards, valuing education and training, promoting effective selection, transition and progression, and defining the outcomes for education and training. They therefore oversee a range of educational standards which are set out in their guide *'Standards for curricula and assessment systems July 2008 and updated 2010*'. For the purpose of this review we were asked to comment more generally on how the MRCGP fulfils the GMC's standards for assessments, more specifically;

1. Whether their standards for assessment for GP specialty training is appropriate to the content and purpose of the curriculum. This standard covers issues such as validity, reliability, feasibility, cost effectiveness and feedback. It also covers requirements that the rationale for the choice of assessment will be documented and evidence based. (Standard 8)
2. Whether assessors/examiners are recruited against criteria for performing the tasks that they undertake. This standard covers issues such as clearly specifying the roles, competencies, experience of assessors and equality and diversity training. (Standard 10).
3. Whether the methods used to set standards for classification of trainees' performance are transparent and in the public domain, and that data about the performance of the test (use of standards, decisions about pass/fail levels, borderline candidates) are described and in the public domain. There are also standards for determining successful completion of CCT, progression and achievement and the right of appeal for certain decisions. (Standard 12)

The GMC also sets standards for the delivery of foundation and specialty training, including GP training, and quality assures the delivery of training against those standards. These standards are set out in their publication *The Trainee Doctor* and apply alongside the *'Standards for curricula and assessment systems July 2008 and updated 2010*' referred to above. *The Trainee Doctor* sets out a series of standards related to postgraduate training and for the purpose of this review, the standards under Domain 3 which cover Equality, Diversity and Opportunity are relevant. They cover compliance with employment law, the Equality Act 2010 and the Human Rights Act. Included in this standard are issues related to monitoring of progress, making reasonable adjustments for disability and for trainees unable to undertake full time training and for training of medical staff in equality and diversity issues.

THE AIMS OF THIS REVIEW.

The primary purpose of this review was to carry out an independent quantitative review of recent MRCGP examination data to establish the extent of failure rates affecting specific groups of doctors, particularly International Medical Graduates and black and minority ethnic UK trained doctors. The review we have carried out is of all the CSA sittings from October 2010-November 2012.

METHODS

The statistical analysis is based primarily on candidates that have taken their first CSA examination between November 2010 and December 2012, which will be referred to as the *main* cohort. For some parts of the report based on outcomes for additional CSA attempts we have added candidates whose first attempt is prior to November 2010, but retook the CSA within the time frame. We were given data on 5744 candidates by the RCGP but were only able to match ethnicity to 5721 candidates.

Four data sets were provided by the GMC. Some of these data sets were derived from data given by the RCGP.

(i) CSA outcome data for the period November 2010 and December 2012.
(ii) AKT outcome data for subjects in the CSA dataset.
(iii) Demographic data and the data from the ARCP data for subjects in the CSA dataset.
(iv) Data for subjects released from RCGP training also referred to as ARCP outcome 4 data.
(v) Data from the GMC on candidates who took the PLAB examination (this included IELTS scores and scores for individual components of the PLAB part 2 exam).

One of the reviewers (AE) also arranged to attend a CSA sitting in May in order to observe the processes, techniques, and training associated with the CSA examination.

STATISTICAL CONSIDERATIONS

For each combination of ethnicity and region of Primary Medical Qualification (PMQ), summary statistics are presented for age, gender and AKT component score. We have then estimated the odds ratio for failure at the CSA. In considering the relationship between outcome in the CSA and (i) region of PMQ and (ii) ethnicity, possible confounding factors are age, gender and clinical knowledge. If age, gender or clinical knowledge correlate with region of PMQ or ethnicity, any difference in outcome could be related to these factors. It is therefore appropriate to obtain estimates of the difference in CSA performance adjusted for these factors as well as unadjusted estimates.

Candidates for the CSA also completed the AKT and so this provides an objective measure of performance that will not be influenced by possible subjective biases regarding region of PMQ or ethnicity. Some candidates had multiple attempts at the AKT and so a choice needed to be made as to which AKT score should be used where there were multiple scores. One might use the candidate's

first score or alternatively one might use that which is closest to the CSA attempt. In the analysis presented in the report we have used the score for the first AKT attempt, but we have also carried out an analysis using the score for the attempt closest to the CSA attempt being considered.

We have then estimated the odds ratio for failure at the CSA comparing every other combination of region of PMQ and ethnicity against white UK candidates adjusted for age, gender and AKT component scores. These have been estimated using logistic regression. As the pass mark for the AKT varies between sittings this is also included as a covariate. We also give the unadjusted estimate as a comparator for the adjusted estimate to illustrate the effect of adjustment. Deanery of first ARCP report was included as a random effect into the model to account for possible clustering of outcome by deanery. The effect of clustering is measured by the intra-cluster correlation (ICC), which is a measure of the proportion of variation that is between units so that an ICC equal to zero implies there is no clustering effect.

RESULTS

Characteristics of sample

Table 1 gives the breakdown of ethnic groups by region of PMQ for the main cohort. These are grouped as White, BME or not known. For the purpose of this report ethnicity refers to the binary classification white and BME. Preliminary analysis suggested an interaction between ethnic group and region of PMQ. It was therefore felt that our analysis needed to consider PMQ region broken down by ethnicity. For this reason 626 candidates without ethnic group coding were excluded from further analysis giving a cohort of 5,095. Table 2 summarises the proportion of BME candidates by PMQ region. Table 3 gives the gender and age breakdown by PMQ region and ethnic group.

Collectively these tables show the ethnic profile, gender, age and region of PMQ for the main cohort. The majority (93%) of International Medical graduates (IMG) are classified as being from a black and minority ethnic group (BME). Thirty two per cent of UK graduates are classified as from BME groups. IMG tend to be older (36 years) than UK graduates (30 years) at the first sitting of the CSA examination. This probably reflects the fact that they have completed a period of medical training in their own countries before coming to the UK.

Table 4 gives the breakdown of PMQ region and ethnic group by Deanery of the first ARCP report. At the time of writing deaneries were still the responsible authorities for postgraduate training in the NHS. Their functions will shortly be taken over by Local Education Training Boards but we have referred to them as deaneries throughout this report. Table 4 shows the distribution of region of training by Postgraduate Deanery for the main cohort. One of the factors that we have to control for is the possibility that the standard of training may vary between deaneries, hence contributing to the differential failure rate that we are investigating. This table shows that the proportion of British BME and IMG varies substantially between different deaneries. So for example the proportion of trainees who are IMG are much greater in the East Midlands, East of England, North Western and Mersey Deaneries compared to London, Oxford, and the west of the country (Wessex, Peninsula, Severn).

Outcome of CSA

Table 5 gives the breakdown of CSA marks by different domains (Data Gathering, Technical and assessment skills, Clinical Management and Interpersonal skills) of the CSA and the total CSA score by PMQ region and ethnic group. Since the pass mark for the CSA varies between sittings (because of the standard error methodology referred to earlier) we also give the mean mark relative to pass mark. For the first attempt the failure rates differ substantially between PMQ regions with BME candidates having a higher rate within each PMQ regions. However, the greatest differences in marks are between UK and non UK graduates.

Table 6 gives the CSA failure rate for the first 4 attempts. For attempts 2, 3, and 4 subjects taking the examination between November 2010 and December 2012, but taking their first CSA before November 2010 are included.

Table 7 gives a breakdown of the CSA failure rate at first attempt by age, ethnic group and PMQ region. Table 8 given the corresponding breakdown by gender. There is evidence that age and gender influence outcome in the CSA irrespective of ethnicity and PMQ region with women and younger candidates having a lower failure rate at the first attempt. Given that non UK candidates tended to be older with a lower proportion being female, outcome according to region is confounded with age and gender.

The Applied Knowledge test (AKT) is a machine marked summative assessment of knowledge that underpins General Practice and broadly speaking measures applied knowledge (the 'knows how' of Miller's pyramid). It has been suggested as a reliable predictor of a candidate's performance in the CSA and therefore a potential confounder. Table 9 gives results for the candidate's first AKT attempt by ethnicity and PMQ region. It shows that both UK BME and non- UK candidates have a lower success rate at their first AKT when compared to white UK graduates. The average mark for the AKT was higher for UK White graduates followed by UK BME graduates. However, the greatest differences are once again between UK and non-UK graduates. Table 10 gives the failure rate at the first CSA according to success at the first AKT, showing that failure at the first AKT increases the failure rate at the first CSA attempt irrespective of ethnicity or PMQ region.

Modelling CSA Pass Rate.

As well as ethnicity and PMQ region, age gender and outcome of the AKT appear to affect the success rate at the CSA. A multivariate analysis is therefore needed to adjust for confounding variables. We have chosen to use a logistic regression model to obtain estimates of the odds ratio of failure between groups adjusted for possible confounding variables. Based on the descriptive analysis listed in Tables 1-10 we believe that it is appropriate to include age, gender and AKT component scores (Clinical Medicine, Evidence Interpretation, and Organisational Issues) as covariates. Some candidates had multiple attempts at the AKT. Separate analyses were carried out using (i) the candidates' AKT scores at the first attempt and (ii) the AKT score closest to the CSA sitting. These gave very similar results so we have presented an analysis using the results of the first AKT for the analysis of CSA outcomes. As the pass mark can change between sittings of the AKT an alternative would have been to use the total mark relative to the pass mark, but this would have

precluded using the component scores. So that component scores could be used, the pass mark was therefore included as a covariate to account for the variable pass mark between sittings.

Inspection of failure rates by deanery suggested that there were some variation in success according to deanery (Table 11). Northern Ireland had the lowest failure rate (2.5%) whilst Kent, Surrey and Sussex had the highest (37.4%). To take account and investigate this source of clustering, deanery of first ARCP report was included as a random effect into the model giving a random effect logistic model[7]. The effect of clustering is measured by the intra-cluster correlation (ICC), which is a measure of the proportion of variation that is between units. An ICC equal to zero implies there is no clustering effect.

Analyses were carried out for the first, second, and third attempts at the CSA, but not the fourth attempt as there was insufficient data in some groups (see Table 6) to fit the model. To illustrate the effect of inclusion of covariates, unadjusted analyses that are without covariates and adjusted analysis are presented in Table 12 and Table 13 respectively.

Model coefficients have been presented as odds ratios. An odds ratio equal to 1 corresponds to no effect. For ethnicity and PMQ region odds ratios greater than 1 imply a higher failure rate for the group compared to white candidates with a UK PMQ. This is also the case for females where the odds ratio of women passing the CSA is compared to men. For the quantitative variables, in this case age and the AKT components, the odds ratio is the increase in the odds for a 1 unit increase in the scale.

CSA First Attempts

Considering first the unadjusted odds ratios (Table 12). All five groups defined by ethnicity and PMQ do significantly worse than White UK graduates at their first attempt. The smallest difference is between UK BME with an odds ratio of 4.8 (95% c.i. 3.7 to 6.1,p<0.001) and the greatest for EEA BME with an odds ratio of 45.7 (95% c.i. 23.9 to 87.4,p<0.001). These odds ratios reflect the substantial difference in pass rate seen in Table 9.

Table 13 gives estimates for the model adjusted for gender, age and AKT score. We can see from the table that Women and candidates with higher scores on their AKT are less likely to fail the CSA at the first attempt consistent with the effects seen in Table 8 for gender and Table 10 for the AKT. Older candidates are less likely to pass consistent with Table 7. All five groups have a significantly higher failure rate compared to white UK graduates, although the odds ratios are now closer to 1 than in Table 12, with an odds ratio of 3.5 (95% c.i. 2.7 to 4.6,p<0.001) for UK BME applicants and the greatest effect is for IMG BME candidate with an odds ratio of 10.1 (95% c.i. 5.0 to 20.4,p<0.001). Using a contrast of model coefficients an adjusted estimate of the odds ratio for BME compared to white for non- UK candidates was obtained. BME candidates were more likely to fail than white candidates (OR=3.8,95% c.i. 1.5 to 9.8,p=0.006).

In both the unadjusted and adjusted analysis a random effect was included for Deanery. There was only slight evidence of clustering with an intra-cluster correlation (ICC) equal to 0.0246 in the unadjusted model reducing to 0.009 in the adjusted model. The reduction in the ICC is to be expected as a covariate will tend to remove difference in the characteristic of candidates between deanery.

CSA Second and Third Attempts

A similar pattern was observed for second and third attempts at the CSA, although now the odds ratios were in most case closer to one (Table 12 & Table 13). Of note within this, the pass rate for UK BME was similar to UK white candidates at a second attempt with unadjusted odds ratio of 1.1 (95% c.i. 0.6 to 2.0,p=0.786) and an adjusted odds ratio of 1.0 (95% c.i 0.5 to 1.9, p=0.968). Intra cluster correlation was negligible in all models for the second and third attempts.

The most important finding demonstrated by Table 13 is that even when controlling for age, gender and AKT component score, all groups do significantly worse than White UK graduates in CSA failure rates. The greatest difference is shown for BME international medical graduates. The differences decrease with subsequent sittings of the CSA, disappearing for BME UK graduates at the second CSA sitting. The differences still persist for BME international medical graduates at the second and third attempts.

IELTS and PLAB Part 2 Scores

As mentioned earlier, in order to obtain full registration by the GMC and prior to entering medical practice non- EEA IMG will usually be expected to complete the IELTS and the PLAB examinations. To further investigate factors that might influence performance in the CSA we carried out analysis of the cohort of candidates who had taken these examinations. Additional data used for this analysis included the IELTS scores, which include components for reading, speaking, understanding and writing. We also included the scores of candidates sitting the PLAB Part 2 which is an OSCE examination not dissimilar to the CSA test. It assesses the competencies of IMG graduates to practise medicine safely in UK hospitals and the standard is set at the level of what would be expected of a trainee completing a Foundation Year One (F1). The components of the test assess clinical examination, practical skills, communication skills and history taking. We did not use outcome data from PLAB Part 1 because we already had data from the AKT which is a similar machine marked test but has the advantage that it is taken shortly before candidates attempt the CSA. In our view the AKT was likely to be a better predictor of CSA then PLAB Part 1.

The GMC is already undertaking a review of the PLAB examination specifically looking at whether international medical graduates granted full registration after passing the PLAB test are more or less likely than other cohorts of doctors to experience difficulties in medical practice in the UK.

Table 14 gives the distribution of Non-UK graduates taking IELTS and/or PLAB Part 2.There is a substantial overlap in the candidates taking both IELTS and PLAB Part 2. Mean outcome for IELTS components and the overall score were similar for all groups and region of PMQ (Table 15). Table 16 gives the failure rate on the first CSA broken down by IELTS scores. Pass rate on the CSA at the first attempt increased with all IELTS component scores. Currently registration requirements for the GMC require an IMG to score 7 on all components of the IELTS and it is worth noting that if the IELTS

requirements were increased, then the failure rate at the CSA would decrease. It appears that the Understanding component is the most important predictor for CSA pass rate.

Considering now the PLAB Part 2 some applicants with Non UK PMQ had been exempted. Table 17 gives the failure rate by ethnicity, PMQ region and PLAB Part 2 exemption. The failure rate is similar for exempt and not exempt BME candidates, whereas white candidates that were exempt had a lower rate than not exempt candidates. Table 18 gives the pass rate for bands of PLAB Part 2 Components. As with IELTS, higher scores on PLAB Part 2 are associated with a higher pass rate in the CSA.

Modelling CSA Outcome in candidates taking IELTS and PLAB Part2

A logistic regression model was fitted with a random effect for deaneries to investigate the association between CSA outcome and IELTS and PLAB Part2. This model is based on data for non-UK candidates. Comparison is between BME and white and between EEA and IMG as there was not an interaction between PMQ and ethnicity. Details of the covariates are the same as for the model in Table 13 with the addition of covariates for IELTS components and PLAB Part 2. Table 19 and Table 20 give the unadjusted and adjusted analysis respectively. In this cohort of IMG and EEA candidates, BME candidates were more likely to fail, but this was no longer statistically significant (Adj. OR=1.6, 95% c.i. 0.88 to 2.8, p=0.127) and the odds ratio was closer to one than the corresponding odds ratio (Adj OR=3.8) in the analysis without adjustment for IELTS and PLAB Part 2 given in the text above derived from Table 13 or the unadjusted analysis in Table 19(OR=2.53,p<0.001).

Other covariates that predicted outcome in this were gender (Adj. OR=0.50) with women less likely to fail, age (Adj. OR=1.1) indicating lower pass rate in older candidates and AKT Clinical medicine component (Adj. OR=0.96), IELTS Understanding Score (Adj. OR=0.72), and PLAB Part 2 Communication (Adj. OR=0.61) all indicating that higher scores in these exams reduced the CSA failure rate at the first attempt.

These analyses show that issues related to what we have termed 'clinical understanding' which is a complicated concept and may include linguistic understanding as well as clinical reasoning and communication and which may be indirectly measured by components of the IELTS, PLAB Part 2 and the AKT are probably related to and predictive of failure rate in the CSA. We are not able to directly compare UK and non UK graduates in this because UK graduates do not sit the IELTS and PLAB examinations. However, the fact that the odds ratio between BME and White non-UK graduates fell from 3.8(Adj, OR =3.8, 95% c.i. 1.5 to 9.8,p=0.006) to 1.6(Adj. OR=1.6, 95% c.i. 0.88 to 2.8, p=0.127) when these components are taken into account suggests that this may be important in understanding why IMG have such a high failure rate in the CSA examination.

Release from RCGP Training Program (ARCP Outcome 4)

When we were asked to explore the failure rate of specific groups of doctors in the CSA, the GMC was also concerned because of reports that many doctors were forced to leave GP speciality training because of failure in the CSA examination. As mentioned previously, the MRCGP is a high stakes examination and failure to pass after 4 attempts means that doctors can no longer work in general practice in any capacity. However, many doctors may leave training for other reasons – for example

failure to pass the work based assessments, not satisfying the Annual Review of Competence Progression (ARCP) or because of other issues of competence. These are classified as' ARCP Outcome 4'.

Data were provided on 374 trainees released from the RCGP training programme with ARCP Outcome 4 of whom 176 match to the main cohort and had data on ethnicity. Table 21 gives the number and percentage released by Ethnicity and PMQ region. The rate of release was substantially higher for candidates with PMQ from outside the UK. Table 22 gives the corresponding information for gender with women less likely to be released. Table 23 compares the characteristics of trainees released from the programme with the remainder of the main cohort. Released trainees were on average older and had lower score on the both the CSA and the AKT. Table 24 give the numbers released by deanery with Mersey showing the highest release rate (11.5%) compared to rate across all deaneries of 3.5%.

To understand the multivariate relationship between these factors a logistic regression model was fitted including ethnicity, PMQ region, gender, age, CSA and AKT score with a random effect for deaneries. In this instance the last set of CSA and AKT scores were used for each candidate. Table 25 gives the unadjusted odds ratios. Table 26 gives the coefficients of the model expressed as adjusted odds ratios. Compared to White UK trainees other groups were more likely to be released with the following odds ration compared to White UK: BME UK (Adj. OR =2.8), White IMG (Adj.OR= 4.3), BME IMG(Adj. OR=8.3) , EU White (Adj. OR =6.6) and EU BME (Adj. OR =13.7). Amongst other covariates gender and CSA interpersonal skills components were the most strongly associated with release with men and lower interpersonal skill being associated with increased risk of being released.

Table 27 gives the distribution of the number of CSA attempts for trainees released from the programme and the outcome for their last attempt. For example 13 trainees that were released took the CSA only once of whom only 2 failed suggesting that outcome 4 did not relate to the CSA in these cases. In contrast 39 trainees took the CSA 4 times of whom 33 failed on the last attempt.

Summary conclusions

Is there a differential outcome for different ethnic groups in the MRCGP examinations?

The AKT Examination

Our results clearly show that there are significant differences in outcome in both the AKT and the CSA components of the MRCGP examinations between UK BME and IMG BME candidates in comparison to White UK graduates.

The AKT is a machine marked examination testing applied clinical knowledge. There is a differential pass rate for both BME UK graduates and IMG graduates when compared to White UK graduates (Table 9). It is difficult to attribute this to bias because of the nature of the test and the reasons for the differential pass rates are likely to be complex.

Differentials between White UK graduates and BME UK graduates seem to reflect existing observed differentials in examination performance which are described both in Higher Education and in Medical examinations[8;9]. There is a general consensus that the reasons for this are complex. The differences in Higher Education where there is a difference between degree outcome and ethnicity have been extensively studied and persist despite controlling for factors such as prior attainment, social class and school background. They have persisted for many decades and we currently do not have clear interventions to reduce these differences.

The biggest differential is between UK graduates and IMG graduates suggesting that it is the preparedness of the candidates based on prior education experience that may be a factor. Within this group we have no information on prior attainment so can only speculate as to the reasons for the differential outcomes. The AKT is an applied knowledge test relevant for UK General Practice. The vast majority of IMG candidates come from the Indian subcontinent and from other countries where the discipline of General Practice is poorly developed. IMG candidates will therefore have much less direct experience of General Practice than their UK counterparts. In our view, this must disadvantage this group in subtle ways and explain the much larger differences in outcomes between UK and non-UK graduates. This will also be one of the reasons that there are significant differences in outcome in the CSA examination.

The CSA Examination

Our results show that there are significant differences in failure rate between different groups in the CSA examination. Even after controlling for age, gender and performance at AKT, significant differences persist between White UK graduates and BME UK graduates (Table 13). BME UK graduates are nearly four times more likely to fail the CSA examination at their first attempt than their White UK colleagues (OR = 3.536, c.i 2.701-4.629, p= <0.001). BME IMG candidates are nearly fifteen times more than likely to fail this exam than their White UK colleagues (OR= 14.741, c.i. 11.397-19.065, p=<0.001).

The differences are much greater between UK and non-UK graduates suggesting that it is the preparedness of UK graduates that may be an explanation for the differences between these two

groups. The CSA is not a culturally neutral examination and nor is it intended to be. It is not and nor should it be just a clinical exam testing clinical knowledge in a very narrow sense. It is designed to ensure that doctors are safe to practise in UK general practice. The cultural norms of what is expected in a consultation will vary from country to country. So for example, a British graduate will have difficulty in practising in a general practice setting in France or in India until they become acculturated to that system of care. British graduates have much greater exposure, both personally and through their training, to general practice when compared to the majority of IMG who graduate from health systems which are not as dominated by primary care as the NHS. Most medic al schools in the UK now have well developed programmes for communication skills training, reflective practice and direct exposure of students to General Practice as a discipline. Approximately 40% of foundation training programmes will require a UK graduate to spend some time in a general practice setting. We have not been able to analyse data on success in CSA based on training experience. However, for those who have gone through foundation training which included time in a GP setting, it does mean that when a UK graduate sits the MRCGP examination they will have had much greater exposure to a general practice setting than most IMG. This could place them at a significant advantage when compared to their IMG colleagues. As the number of foundation programmes that include time in a GP setting increases, it may be that disparities between IMG and UK graduates in CSA outcomes will increase.

The nature of the examination is such that it is open to subjective bias. We cannot ascertain if the standardised patients (played by actors) behave differently in front of candidates from non-White ethnic groups. Nor can we confidently exclude bias from the examiners in the way that they assess non-White candidates. However, having observed (by AE) the examination and read the background documentation, it is clear to us that the RCGP is aware of these potential biases and takes steps to mitigate them. So for example there is mandatory training of RCGP examiners in equality and diversity issues, there is training and monitoring of the actors to ensure consistency in the presentation of the cases, and there is a well-developed programme of continuing training and feedback to examiners of their performance.

It is also our view that the method of assessment is not a reason for the differential outcomes that we have described. The CSA examination and the marking of the exam is based on a well-established pedagogy which is internationally recognised and used widely in postgraduate examinations[10]. This includes the borderline group method of setting the standard in the CSA. There is controversy about the use of standard error measurements to create the 'cut score' but it is beyond the scope of this review to comment on this. However, like any clinical examination, the CSA is subject to bias and there are areas where its delivery could be improved. The RCGP itself has been at the forefront of research to understand the biases caused by oral examinations[11].

Some people have argued that the fact that candidates seem to have a lower failure rate in the work placed based assessments (WPBA) suggests that the CSA is flawed as an assessment method. However, it is important to recognise that the CSA is just one component of the assessment of general practice trainees – it is testing different skills and knowledge when compared to the WPBA and so cannot be directly compared.

It is also worth pointing out that in our observations of the CSA examination, we noted that the pool of examiners was not representative of the ethnic background of general practitioners more generally. Our understanding is that the RCGP does attempt to recruit examiners from a diverse range of backgrounds and acknowledges that the current pool of examiners is not representative of GPs in the country. Part of the reason could be due to the criteria that they have established for the selection and recruitment of examiners. These should be reassessed if they continue to fail to recruit a diverse pool of examiners. This is not just about tokenism and making sure that the numbers meet some arbitrary diversity target but a recognition that encouraging a diversity of views and opinions amongst the examiners will contribute to changing the perception of the examination for candidates, will challenge in a positive manner some of the standards that are set for assessing the candidates and be more reflective of the nature of general practitioners in the country.

We also noted that on the day that we observed the examination that there were very few actors representative of ethnic minority backgrounds and that there were no cases that assessed the ability of candidates to consult across cultures. There may well be a series of cases that assess this in the totality of examination cases that they have developed. We simply make the point that depending on where you practise as a GP, you are likely to see a huge difference in the diversity of patients and the problems that they present with. It is important that the candidates are assessed in a way that reflects the diversity of patients that they see. The type of cases which present in our major conurbations to general practitioners where the population is ethnically diverse are very different from the presentation of cases in areas where there is less ethnic diversity and it may be that the current examination does not reflect this diversity in the cases that are chosen for examination.

We also observed that the feedback given to candidates was limited. We were told that the number of candidates precluded individualised feedback. We feel that this is not acceptable especially in an examination that charges a high fee and which is a high stakes exam such as the MRCGP. If candidates fail the exam they need to know why, through a process of formative feedback to both the candidate and their trainer. Mechanisms should be developed to enable this to happen. It is interesting to note that differences between White UK and BME UK graduates disappear at the second attempt of the CSA and also reduces for IMG candidates (Table 13). This may reflect on the feedback and better preparation but our comments regarding feedback are still pertinent.

As pointed out earlier, the largest differences in pass rates are between UK and non-UK graduates. Whilst observing the examination, we noted that the weaker candidates who were failing the clinical stations appeared to be less well prepared than the candidates who were doing well. This could be a reflection of the training they were receiving in their workplace or the fact that they had been less exposed to training in general practice because they did not graduate in this country. So long as this country depends on recruiting large numbers of international medical graduates, then we need to acknowledge that most of these IMG come into medicine from a different starting point. Many will require much more training and support before they can be considered equivalent to their British colleagues and perhaps the differential outcome in the CSA examination is a reflection of this. Our observations suggest that IMG are treated exactly the same as British graduates, perhaps through a misguided attempt at being fair whereas what is needed is an explicit acknowledgement of the problems that they are likely to face with advice on how these can be mitigated. We are aware that many deaneries provide additional help to IMG (through exam preparation courses) but the level of support needs to go beyond this and may require fundamental changes to the structure of the

training programmes. For example there may be a case for spending the existing hospital based part of the programme in posts /specialties that would give IMG greater opportunities to increase their communication and clinical reasoning skills. Another area for improvement could be in relation to the information that IMG are given. So for example, whilst there is generic advice available through the RCGP on characteristics of candidates who fail the exam, there is no explicit acknowledgement of the problems that IMG may face nor advice on how to improve their chances. In *The Trainee Doctor* publication, the GMC seems to acknowledge that women who take family leave and disabled candidates may need additional support and bespoke training. Perhaps there needs to be an acknowledgment of the additional needs of some IMG doctors.

If we are willing to accept that it is the lack of familiarity with general practice and the context of the MRCGP examination that accounts for the differential performance of IMG in the CSA examination then we need to develop interventions that can address these deficiencies.

Are there differences in outcome based on Deanery of training?

There is clearly a large difference in place of training for doctors with some deaneries having a very large proportion of their trainees being IMG (Table 4). This should not be regarded as a marker of relative training quality and there needs to be further work to look at the demographics of those entering training and their scores at the point of entry compared with their performance in the CSA. We have no information on the quality of training in these deaneries but the combination of selection and training placement systems may operate against the interests of the weaker recruits – in this case IMG. What this means in practice is that those candidates performing least well at selection are assigned to the least popular training placements, thereby encouraging a cycle of educational deprivation[12]. Seeking to counter this systematic unintended discrimination could be the single most important way of ensuring the highest standards of training. Although our analyses did not show clustering of outcome by deanery this does not apply for those trainees who were released from training (Table 26), suggesting that the place of training is important.

Perhaps deaneries where there are a large proportion of IMG should explicitly acknowledge that this group might need additional training and support and place the candidates in their stronger training practices. We are aware that this is happening in some deaneries and it may be that in order to avoid accusations of bias towards some candidates and stigmatisation of IMG, these deaneries could provide extra training opportunities for all their trainees, recognising that it will differentially help and support IMG. Rather than expect all deaneries to adhere to some national standard of training, the GMC should perhaps insist that deaneries with a high proportion of IMG put in place additional support mechanism for their trainees.

Do doctors fail to complete their training because of failure to pass the MRCGP examination?

ARCP outcomes

It seems that most candidates whether UK BME or IMG do ultimately pass the AKT examination. However this is not the case for the CSA examination with nearly 142 BME-IMG candidates from our main cohort being released from the training programme. We are not clear as to the exact reasons for this since only 53 candidates had taken the CSA examination more than four times (Table 27). Candidates leave the programme not having used up all their attempts and our estimate is that 100 candidates get an 'ARCP Outcome 4' before their fourth and final attempt at the CSA. This may be due to reaching the maximum number of attempts for AKT or the maximum training time extension. We do not think that this data set is complete and it is important that more detailed information is collected on this group.

Recommendations:

1) There should be continued monitoring of outcomes in the AKT and CSA examinations with all candidates being aware of the outcomes by different ethnic groups. There should be clear guidance on a framework for monitoring the outcome of high stakes examinations so candidates are aware of the outcomes and regulators are aware of significant deviations in patterns of success and failure.

2) IMG should be made explicitly aware of the differential outcomes with clear advice on how to better prepare for the examination. The current GMC website has specific advice for IMG sitting the PLAB examination and the RCGP should consider providing similar advice for IMG with clear advice on training for those who may not have had sufficient exposure to general practice during their undergraduate and postgraduate training. Candidates need to be made aware of the relationship between IELTS, PLAB scores and the outcome of the CSA examination so that they can focus on improving the areas that they are weak on.

3) As part of the standard setting process for the CSA, the GMC should pay particular attention to the diversity of examiners for the MRCGP, the case mix of exam stations ensuring that they reflect the norms of general practice in a multi-cultural society, the training of standardised patients (including equality and diversity training) and the diversity of the standardised patients. Further research should be commissioned, by the GMC to investigate how BME standardised patients and BME examiners score candidate physicians who are racially and ethnically concordant and compare that to how non concordant standardised patients and examiners score the BME candidates.

4) The GMC should also develop clear guidelines on an acceptable format for formative feedback which will give all candidates clear advice on their areas of weakness and how these can be addressed.

5) Consideration should be given to developing additional training standards for deaneries/LETBs where there are a large proportion of IMG trainees. There needs to be a clear recognition that training programmes need to take account of the fact that doctors are entering training from different starting points and that some trainees may need to have more tailored support. This should include training for educational supervisors and trainers who need to be aware of the differential outcomes for certain groups of trainees and develop appropriate interventions. We do not know if this additional support will improve outcomes and the benefits of any interventions should be appropriately evaluated. Consideration should also be given to commissioning research to assess outcome in the

MRCGP examination by training route and the impact of exposure to a GP setting in foundation training on pass rates for candidates.

6) There should be better linkage of assessment data throughout training including PLAB/IELTS and recruitment data from the NRO which can be fed into training programmes enabling GP trainers to have a better understanding of the strengths and weaknesses of the trainees that they will be supervising. This may enable them to develop individualised and appropriate interventions. A huge amount of data is already collected by the different bodies responsible for training and recruitment and this should be integrated and used for enhancing training support and not just for monitoring. Together with the appropriate level of support from their trainers and educational supervisors in the deaneries/LETB, trainees would be better prepared to sit examinations and potentially have better outcomes.

7) Data from the selection scores of doctors recruited into general practice and held by the NRO should be integrated with CSA outcome data so that we can better understand the relationship between attainment at this level and CSA outcome. This will reinforce the case for more targeted support for weaker candidates that we appear to have identified. The advantage of this data set is that it can be used for both UK and non-UK graduates.

8) There should be better linkage between Foundation assessments, PLAB, IELTS, ARCP data and Examinations data by the GMC. It is important to understand exactly how many candidates leave training because of failure at the CSA examination, especially for those candidates who have used up all their attempts. Currently the data sets are not held by one organisation nor are they all robust enough to assess this information.

9) The Deaneries/LETBs need to have clear information available as to the exact reason that trainees leave training programmes. The GMC should insist on this information being available to them as part of their regulatory functions. Exit interviews with clearly recorded outcomes may be the best way of collecting this information. It is important because failure to complete a training programme represents a significant loss both to the individual, the profession and the country.

10) The GMC should commission more research on understanding why women consistently outperform men and on IMG who pass the different MRCGP examinations. What are the traits, learning styles and examination techniques that make these candidates succeed? It is better to focus on reasons for success rather than understanding failure because this may suggest ways in which apparent barriers to success may be overcome.

Table 1 Gender and Age by Ethnic Minority by Region of Primary Qualification

Ethnic Group	UK	IMG	EEA	Total
White				
English/Welsh/Scottish/N. Irish	2,322 (55.9)	17 (1.2)	10 (6.3)	2,349 (40.9)
Republic of Ireland	73 (1.8)	(0.0)	13 (8.2)	86 (1.5)
Other	89 (2.1)	69 (4.8)	70 (44.0)	228 (4.0)
Sub Total	2,484 (59.8)	86 (6.0)	93 (58.5)	2,663 (46.4)
BME				
White and Black Caribbean	3 (0.1)	1 (0.1)	(0.0)	4 (0.1)
White and Black African	10 (0.2)	15 (1.0)	2 (1.3)	27 (0.5)
White and Asian	42 (1.0)	14 (1.0)	(0.0)	56 (1.0)
Other Multiple Ethnic Background	33 (0.8)	9 (0.6)	1 (0.6)	43 (0.7)
Asian/Asian British - Indian	462 (11.1)	536 (37.4)	16 (10.1)	1,014 (17.7)
Asian/Asian British - Pakistani	216 (5.2)	302 (21.1)	7 (4.4)	525 (9.1)
Asian/Asian British - Bangladeshi	47 (1.1)	23 (1.6)	4 (2.5)	74 (1.3)
Asian/Asian British - Chinese	73 (1.8)	14 (1.0)	(0.0)	87 (1.5)
Asian/Asian British - Other	149 (3.6)	98 (6.8)	6 (3.8)	253 (4.4)
Black/Black British - African	53 (1.3)	140 (9.8)	6 (3.8)	199 (3.5)
Black/Black British - Caribbean	6 (0.1)	8 (0.6)	1 (0.6)	15 (0.3)
Black/Black British - Other	9 (0.2)	8 (0.6)	(0.0)	17 (0.3)
Arab	3 (0.1)	4 (0.3)	(0.0)	7 (0.1)
Other ethnic group	54 (1.3)	52 (3.6)	5 (3.1)	111 (1.9)
Sub Total	1160 (27.9)	1224 (85.4)	48 (30.2)	2432 (42.3)
Not Known				
Prefer not to say	61 (1.5)	19 (1.3)	3 (1.9)	83 (1.4)
Missing	446 (10.7)	105 (7.3)	15 (9.4)	543 (9.5)
Sub Total	507 (12.2)	124 (8.6)	18 (11.3)	626 (10.9)
Total	4,151	1,434	159	5,744

Table 2 Ethnic Minority by Region of Primary Qualification used for analysis

	Region of Primary Qualification			
	UK	IMG	EEA	Total
BME	1,160	1,224	48	2,432
(%)	(31.8)	(93.4)	(34.0)	(47.7)
Total	3,644	1,310	141	5,095

Table 3 Gender and Age by Ethnic Minority by Region of Primary Qualification

	Region	UK		IMG		EEA		
	Ethnic Grp	White	BME	White	BME	White	BME	Total
Gender	Female	1,693	647	54	511	60	19	2,984
	(%)	(68.2)	(55.8)	(62.8)	(41.8)	(64.5)	(39.6)	(58.6)
Age at Time of First CSA (years)	<30	1,330	646	3	17	7	4	2,007
	(%)	(53.5)	(55.7)	(3.5)	(1.4)	(7.5)	(8.3)	(39.4)
	30-34	882	432	27	477	48	21	1,887
	(%)	(35.5)	(37.2)	(31.4)	(39.0)	(51.6)	(43.8)	(37.0)
	>=35	272	82	56	730	38	23	1,201
	(%)	(11.0)	(7.1)	(65.1)	(59.6)	(40.9)	(47.9)	(23.6)
	Mean	30.5	30.1	37.0	36.4	34.3	35.3	32.0
	Median	29	29	36	35	34	34	31
	5th Centile	27	27	30	31	29	29	27
	95th Centile	38	35	45	45	41	47	41
	N	2484	1160	86	1224	93	48	5095

Table 4 First Deanery by Ethnic Group by Region of Primary Medical Qualification

	UK				IMG				EEA				Total
	White		BME		White		BME		White		BME		
	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n
Defence	36	(94.7)	2	(5.3)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	38
East Midlands	111	(38.0)	64	(21.9)	6	(2.1)	104	(35.6)	4	(1.4)	3	(1.0)	292
East of England	122	(31.8)	83	(21.6)	7	(1.8)	157	(40.9)	9	(2.3)	6	(1.6)	384
Kent, Surrey and Suss	145	(29.8)	117	(24.0)	16	(3.3)	184	(37.8)	16	(3.3)	9	(1.8)	487
London	267	(42.9)	298	(47.9)	10	(1.6)	39	(6.3)	4	(0.6)	4	(0.6)	622
Mersey	116	(53.2)	29	(13.3)	3	(1.4)	61	(28.0)	5	(2.3)	4	(1.8)	218
Scotland (East)	32	(72.7)	4	(9.1)	2	(4.5)	5	(11.4)	1	(2.3)	0	(0.0)	44
Scotland (North)	44	(67.7)	4	(6.2)	4	(6.2)	12	(18.5)	1	(1.5)	0	(0.0)	65
Scotland (South East)	63	(71.6)	11	(12.5)	0	(0.0)	12	(13.6)	2	(2.3)	0	(0.0)	88
Scotland (West)	137	(70.6)	20	(10.3)	2	(1.0)	32	(16.5)	3	(1.5)	0	(0.0)	194
Oxford	106	(59.9)	51	(28.8)	1	(0.6)	17	(9.6)	2	(1.1)	0	(0.0)	177
Wessex	114	(61.6)	29	(15.7)	8	(4.3)	27	(14.6)	4	(2.2)	3	(1.6)	185
Peninsula	85	(80.2)	5	(4.7)	4	(3.8)	9	(8.5)	3	(2.8)	0	(0.0)	106
Severn	155	(77.5)	20	(10.0)	3	(1.5)	17	(8.5)	4	(2.0)	1	(0.5)	200
West Midlands	178	(32.4)	163	(29.7)	6	(1.1)	187	(34.1)	7	(1.3)	8	(1.5)	549
North Western	159	(39.6)	115	(28.6)	5	(1.2)	113	(28.1)	5	(1.2)	5	(1.2)	402
Northern	137	(59.6)	24	(10.4)	4	(1.7)	53	(23.0)	10	(4.3)	2	(0.9)	230
Northern Ireland	95	(93.1)	1	(1.0)	0	(0.0)	1	(1.0)	5	(4.9)	0	(0.0)	102
Wales	116	(60.7)	22	(11.5)	1	(0.5)	48	(25.1)	3	(1.6)	1	(0.5)	191
Yorkshire & Humber	238	(54.1)	73	(16.6)	4	(0.9)	118	(26.8)	5	(1.1)	2	(0.5)	440
Not Specified	28	(34.6)	25	(30.9)	0	(0.0)	28	(34.6)	0	(0.0)	0	(0.0)	81
Total	2,484	(48.8)	1,160	(22.8)	86	(1.7)	1,224	(24.0)	93	(1.8)	48	(0.9)	5,095

Table 5 **CSA Marks for Domains and totals for first attempt by Ethnic Minority by Region of Primary Qualification**

	Region	UK		IMG		EEA		
	Ethnic Group	White	BME	White	BME	White	BME	Total
Data gathering, Technical and assessment skills	mean	30.44	29.02	26.53	24.89	26.96	23.94	28.60
	(sd)	(3.3)	(3.6)	(4.2)	(3.6)	(3.6)	(4.7)	(4.2)
Clinical management skills	mean	27.95	25.59	23.19	21.26	24.03	20.79	25.58
	(sd)	(3.6)	(4.0)	(4.7)	(4.1)	(4.0)	(4.8)	(4.7)
Interpersonal Skills	mean	30.79	28.49	24.67	22.31	25.82	22.50	27.96
	(sd)	(3.5)	(4.2)	(5.6)	(4.4)	(4.2)	(5.6)	(5.3)
Total Score	mean	89.18	83.10	74.40	68.46	76.81	67.23	82.14
	(sd)	(9.0)	(10.6)	(13.3)	(10.9)	(10.6)	(14.2)	(13.2)
Total Relative to Pass Mark	mean	15.58	9.47	0.97	-5.07	3.34	-6.29	8.55
	(sd)	(9.1)	(10.6)	(13.4)	(10.9)	(10.6)	(14.1)	(13.2)
	N	2,484	1,160	86	1,224	93	48	5,095

Note: not all candidates who fail at first attempt re-take the CSA e.g. some run out of AKT attempts

Table 6 CSA Failure Rate for First 4 attempts by Ethnic Minority by Region of Primary Qualification

	Region	UK		IMG		EEA		
	Ethnicity	White	BME	White	BME	White	BME	Total
First Attempt	Fail	111	198	41	798	30	33	1,211
	(%)	(4.5)	(17.1)	(47.7)	(65.2)	(32.3)	(68.8)	(23.8)
	N	2,484	1,160	86	1,224	93	48	5,095
Second Attempt	Fail	40	73	21	387	9	23	553
	(%)	(36.0)	(38.8)	(52.5)	(48.9)	(34.6)	(74.2)	(46.5)
	N	111	188	40	792	26	31	1,188
Third Attempt	Fail	11	14	8	234	4	11	282
	(%)	(55.0)	(45.2)	(47.1)	(53.5)	(36.4)	(68.8)	(53.0)
	N	20	31	17	437	11	16	532
Fourth Attempt	Fail	0	5	4	119	0	3	131
	(%)	(0.0)	(45.5)	(50.0)	(54.8)	(0.0)	(42.9)	(53.3)
	N	1	11	8	217	2	7	246

Table 7 CSA Failure Rate for First attempts by Age at first CSA, Ethnicity and PMQ Region

AGE at First CSA		UK		IMG		EEA		Total
		White	BME	White	BME	White	BME	
<30	Fail	33	77	1	7	0	1	119
	(%)	(2.5)	(11.9)	(33.3)	(41.2)	(0.0)	(25.0)	(5.9)
	N	1,330	646	3	17	7	4	2,007
30-34	Fail	49	89	14	252	18	13	435
	(%)	(5.6)	(20.6)	(51.9)	(52.8)	(37.5)	(61.9)	(23.1)
	N	882	432	27	477	48	21	1887
>=35	Fail	29	32	26	539	12	19	657
	(%)	(10.7)	(39.0)	(46.4)	(73.8)	(31.6)	(82.6)	(54.7)
	N	272	82	56	730	38	23	1,201

Table 8 CSA Failure Rate for First attempts by Gender Ethnicity and PMQ Region

Gender		UK		IMG		EEA		Total
		White	BME	White	BME	White	BME	
Male	Fail	54	132	20	532	13	22	773
	(%)	(6.8)	(25.7)	(62.5)	(74.6)	(39.4)	(75.9)	(36.6)
	N	791	513	32	713	33	29	2,111
Female	Fail	57	66	21	266	17	11	438
	(%)	(3.4)	(10.2)	(38.9)	(52.1)	(28.3)	(57.9)	(14.7)
	N	1,693	647	54	511	60	19	2,984

Table 9 **AKT Marks for Domains and Totals for First attempt by Ethnic Minority by Region of Primary Qualification**

	Region	UK		IMG		EEA		
	Ethnic Group	White	BME	White	BME	White	BME	Total
Clinical Medicine	mean	78.59	75.00	70.29	70.58	73.22	62.45	75.46
	(sd)	(7.8)	(8.6)	(9.0)	(9.1)	(9.1)	(10.8)	(9.1)
Evidence Interpretation	mean	79.47	74.94	63.47	61.66	70.79	56.67	73.52
	(sd)	(12.4)	(13.0)	(17.1)	(15.3)	(14.4)	(15.4)	(15.4)
Organisational	mean	77.24	71.55	65.70	63.77	70.25	58.65	72.21
	(sd)	(11.07)	(11.69)	(13.22)	(12.75)	(12.85)	(14.65)	(13.03)
Relative to	mean	20.7	12.8	1.7	1.7	8.9	-13.8	13.5
Pass Mark	(sd)	(14.81)	(16.18)	(18.16)	(17.38)	(17.40)	(19.41)	(17.94)
Pass Rate	freq	2,257	916	51	707	66	14	4,011
	(%)	(90.9)	(79.0)	(59.3)	(57.8)	(71.0)	(29.2)	(78.7)
AKT Failure Rate	freq	226	244	35	517	27	34	1,083
	(%)	(9.1)	(21.0)	(40.7)	(42.2)	(29.0)	(70.8)	(21.3)
Total	N	2,483	1,160	86	1,224	93	48	5,094

Table 10 **CSA Failure Rate for First attempts by Ethnicity and PMQ Region and Result of first AKT attempt**

AKT First Attempt		UK		IMG		EEA		
		White	BME	White	BME	White	BME	Total
Pass	Fail	68	121	18	410	18	7	642
	(%)	(3.0)	(13.2)	(35.3)	(58.0)	(27.3)	(50.0)	(16.0)
	N	2,257	916	51	707	66	14	4,011
Fail	Fail	43	77	23	388	12	26	569
	(%)	(19.0)	(31.6)	(65.7)	(75.0)	(44.4)	(76.5)	(52.5)
	N	226	244	35	517	27	34	1,083

Table 11 CSA Failure Rate by Deanery

Deanery	White-UK			BME-UK			White-IMG			BME-IMG			White-EEA			BME-EEA			Total		
	Freq	(%)	N	Freq	(%)	N	Freq	(%)	N	Freq	(%)	N	Freq	(%)	N	Freq	(%)	N	Freq	(%)	N
Defence	4	(11.1)	36	0	(0.0)	2	-	-	-	-	-	-	-	-	-	-	-	-	4	(10.5)	38
East Midlands	1	(0.9)	111	12	(18.8)	64	2	(33.3)	6	70	(67.3)	104	1	(25.0)	4	3	(100.0)	3	89	(30.5)	292
East of England	9	(7.4)	122	14	(16.9)	83	2	(28.6)	7	99	(63.1)	157	4	(44.4)	9	0	(0.0)	6	128	(33.3)	384
Kent, Surrey and Suss	7	(4.8)	145	31	(26.5)	117	10	(62.5)	16	125	(67.9)	184	2	(12.5)	16	7	(77.8)	9	182	(37.4)	487
London	10	(3.7)	267	30	(10.1)	298	2	(20.0)	10	18	(46.2)	39	1	(25.0)	4	3	(75.0)	4	64	(10.3)	622
Mersey	8	(6.9)	116	12	(41.4)	29	2	(66.7)	3	48	(78.7)	61	4	(80.0)	5	4	(100.0)	4	78	(35.8)	218
Scotland (East)	3	(9.4)	32	0	(0.0)	4	2	(100.0)	2	4	(80.0)	5	0	(0.0)	1	-	-	-	9	(20.5)	44
Scotland (North)	3	(6.8)	44	1	(25.0)	4	0	(0.0)	4	7	(58.3)	12	1	(100.0)	1	-	-	-	12	(18.5)	65
Scotland (South East)	1	(1.6)	63	1	(9.1)	11	-	-	-	7	(58.3)	12	0	(0.0)	2	-	-	-	9	(10.2)	88
Scotland (West)	7	(5.1)	137	5	(25.0)	20	1	(50.0)	2	21	(65.6)	32	2	(66.7)	3	-	-	-	36	(18.6)	194
Oxford	2	(1.9)	106	8	(15.7)	51	1	(100.0)	1	10	(58.8)	17	0	(0.0)	2	-	-	-	21	(11.9)	177
Wessex	4	(3.5)	114	9	(31.0)	29	5	(62.5)	8	19	(70.4)	27	2	(50.0)	4	2	(66.7)	3	41	(22.2)	185
Peninsula	4	(4.7)	85	0	(0.0)	5	2	(50.0)	4	5	(55.6)	9	0	(0.0)	3	-	-	-	11	(10.4)	106
Severn	7	(4.5)	155	2	(10.0)	20	0	(0.0)	3	10	(58.8)	17	0	(0.0)	4	0	(0.0)	1	19	(9.5)	200
West Midlands	5	(2.8)	178	32	(19.6)	163	4	(66.7)	6	123	(65.8)	187	2	(28.6)	7	6	(75.0)	8	172	(31.3)	549
North Western	7	(4.4)	159	19	(16.5)	115	3	(60.0)	5	68	(60.2)	113	3	(60.0)	5	3	(60.0)	5	103	(25.6)	402
Northern	11	(8.0)	137	5	(20.8)	24	2	(50.0)	4	38	(71.7)	53	2	(20.0)	10	2	(100.0)	2	60	(26.1)	230
Northern Ireland	1	(1.1)	95	0	(0.0)	1	-	-	-	0	(0.0)	1	1	(20.0)	5	-	-	-	2	(2.0)	102
Wales	7	(6.0)	116	2	(9.1)	22	1	(100.0)	1	32	(66.7)	48	1	(33.3)	3	1	(100.0)	1	44	(23.0)	191
Yorkshire & Humber	7	(2.9)	238	9	(12.3)	73	2	(50.0)	4	77	(65.3)	118	4	(80.0)	5	2	(100.0)	2	101	(23.0)	440
Not Specified	3	(10.7)	28	6	(24.0)	25	-	-	-	17	(60.7)	28	-	-	-	-	-	-	26	(32.1)	81
Total	111	(4.5)	2484	198	(17.1)	1160	41	(47.7)	86	798	(65.2)	1224	30	(32.3)	93	33	(68.8)	48	1211	(23.8)	5095

Table 12 Coefficient for Logistic Regression Models of CSA failure rate giving unadjusted estimates of odds ratios by attempt

	First Attempt (N=5095)				Second Attempt (N=1188)				Third Attempt (N=533)			
Group	OR	95%	c.i.	p	OR	95%	c.i.	p	OR	95%	c.i.	p
BME UK	4.776	(3.709	,6.148)	<0.001	1.086	(0.600	,1.966)	0.786	3.876	(0.941	,15.965)	0.061
White IMG	19.432	(12.134	,31.117)	<0.001	5.798	(2.641	,12.731)	<0.001	6.373	(1.348	,30.129)	0.019
BME IMG	39.080	(31.022	,49.232)	<0.001	6.735	(4.101	,11.059)	<0.001	11.769	(3.394	,40.810)	<0.001
White EEA	9.753	(6.037	,15.756)	<0.001	3.673	(1.485	,9.086)	0.005	3.237	(0.570	,18.378)	0.185
BME EEA	45.732	(23.938	,87.368)	<0.001	6.786	(2.858	,16.113)	<0.001	12.462	(2.467	,62.954)	0.002
Constant	0.047	(0.037	,0.059)	<0.001	0.233	(0.145	,0.375)	<0.001	0.177	(0.052	,0.602)	0.006
ICC	0.0246				<0.001				<0.001			

Table 13 Coefficient for Logistic Regression Models of CSA failure rate estimates of odds ratios for first 3 attempts at the CSA adjusted for gender age and AKT score.

	First Attempt (N=5094)				Second Attempt (N=1188)				Third Attempt (N=532)			
Adjusted	Adj. OR	95%	c.i.	p	Adj. OR	95%	c.i.	p	Adj. OR	95%	c.i.	p
Group												
BME UK	3.536	(2.701	,4.629)	<0.001	1.013	(0.537	,1.912)	0.968	5.080	(1.144	,22.566)	0.033
White IMG	7.171	(4.246	,12.110)	<0.001	3.693	(1.593	,8.563)	0.002	7.643	(1.492	,39.156)	0.015
BME IMG	14.741	(11.397	,19.065)	<0.001	4.380	(2.561	,7.491)	<0.001	11.406	(3.068	,42.403)	<0.001
White EEA	5.540	(3.296	,9.313)	<0.001	3.475	(1.331	,9.069)	0.011	4.663	(0.760	,28.621)	0.096
BME EEA	10.144	(5.040	,20.419)	<0.001	2.858	(1.143	,7.143)	0.025	10.888	(1.988	,59.615)	0.006
Female	0.446	(0.374	,0.532)	<0.001	0.433	(0.329	,0.570)	<0.001	0.503	(0.328	,0.770)	0.002
Age at CSA Examination	1.088	(1.066	,1.111)	<0.001	1.063	(1.033	,1.093)	<0.001	1.087	(1.041	,1.136)	<0.001
AKT												
Clin. Med.	0.990	(0.984	,0.997)	0.006	0.995	(0.986	,1.005)	0.348	1.001	(0.986	,1.016)	0.898
Evid. Inter	0.986	(0.978	,0.994)	0.001	0.976	(0.964	,0.987)	<0.001	1.004	(0.987	,1.021)	0.659
Org. Issues	1.028	(0.994	,1.063)	0.114	1.002	(0.956	,1.050)	0.937	0.980	(0.925	,1.037)	0.486
Mark to pass	0.048	(0.000	,4.854)	0.197	2.789	(0.005	,1665.8)	0.753	2.481	(0.001	,6226.7)	0.82
Constant	0.990	(0.984	,0.997)	0.006	0.995	(0.986	,1.005)	0.348	1.001	(0.986	,1.016)	0.898
ICC	0.009				<0.001				<0.001			

Table 14 IELTS and PLAB Part 2 for non UK PMQ candidates

		IMG		EEA		
Exam		White	BME	White	BME	Total
Neither	Freq	21	102	80	33	236
	(%)	(24.4)	(8.3)	(86.0)	(68.8)	(16.3)
IELT	Freq	1	5	2		8
	(%)	(1.2)	(0.4)	(2.2)	(0.0)	(0.6)
PLAB Part 2	Freq	5	31	2	2	40
	(%)	(5.8)	(2.5)	(2.2)	(4.2)	(2.8)
Both	Freq	59	1,086	9	13	1,167
	(%)	(68.6)	(88.7)	(9.7)	(27.1)	(80.4)
Total	N	86	1,224	93	48	1,451

Table 15 IELTS score by component by Ethnicity by Region of Primary Qualification

Region		IMG		EEA		
Ethnic Group		White	BME	White	BME	Total
Reading	mean	7.33	7.23	7.55	7.04	7.24
	(sd)	(0.67)	(0.76)	(1.06)	(0.63)	(0.75)
Speaking	mean	7.52	7.51	7.82	8.15	7.52
	(sd)	(0.70)	(0.65)	(0.75)	(0.90)	(0.66)
Understanding	mean	7.48	7.49	7.41	7.73	7.49
	(sd)	(0.70)	(0.75)	(0.92)	(0.93)	(0.75)
Writing	mean	6.78	7.15	7.27	6.88	7.13
	(sd)	(0.69)	(0.68)	(0.79)	(1.00)	(0.69)
Overall	mean	7.35	7.42	7.55	7.54	7.42
	(sd)	(0.52)	(0.46)	(0.65)	(0.56)	(0.47)
	n	60	1091	11	13	1175

Table 16 **CSA Failure Rate at First Attempt for each component of IELTS Score**

IELTS Component		6	6.5	7	7.5	8	8.5	9	Total
Reading	Fail	84	153	209	171	81	38	19	755
	(%)	(77.78)	(72.51)	(66.77)	(62.87)	(57.45)	(42.22)	(47.50)	(64.26)
	N	108	211	313	272	141	90	40	1,175
Speaking	Fail	-	-	454	1	249	-	51	755
	(%)	-	-	(68.27)	(33.33)	(62.25)	-	(47.66)	(64.26)
	N	-	-	665	3	400	-	107	1,175
Understanding	Fail	38	91	229	192	112	54	38	754
	(%)	(76.00)	(84.26)	(72.24)	(59.63)	(61.54)	(54.00)	(40.00)	(64.22)
	N	50	108	317	322	182	100	95	1,174
Writing	Fail	133	1	425	0	187	-	9	755
	(%)	(69.63)	(50.00)	(65.18)	(0.00)	(60.71)	-	(45.00)	(64.26)
	N	191	2	652	2	308	-	20	1,175
Overall Score	Fail	-	-	380	257	94	21	3	755
	(%)	-	-	(71.97)	(67.28)	(47.47)	(37.50)	(27.27)	(64.26)
	N	-	-	528	382	198	56	11	1,175

Table 17 **CSA Failure Rate by Region and Exemption from PLAB part 2**

Exemption		White-IMG	BME-IMG	White-EEA	BME-EEA	Total
Yes	Fail	7	72	24	23	126
	(%)	(31.82)	(67.29)	(29.27)	(69.70)	(51.64)
	N	22	107	82	33	244
No	Fail	34	726	6	10	776
	(%)	(53.13)	(65.00)	(54.55)	(66.67)	(64.29)
	N	64	1,117	11	15	1,207

Table 18 **CSA Failure Rate at First Attempt by outcome for each component of first PLAB Attempt Component**

PLAB Part 2 * Component		<3.0	3.0-	3.5-	≥4.0	Total
Communication	Fail	320	318	119	19	776
	(%)	(73.56)	(64.24)	(52.89)	(36.54)	(64.29)
	N	435	495	225	52	1,207
Examination	Fail	307	277	167	25	776
	(%)	(72.4)	(61.8)	(58.4)	(51.0)	(64.3)
	N	424	448	286	49	1,207
History	Fail	306	330	126	14	776
	(%)	(74.1)	(63.6)	(53.4)	(35.9)	(64.3)
	N	413	519	236	39	1,207
Practical	Fail	196	240	231	109	776
	(%)	(69.3)	(67.4)	(62.1)	(55.6)	(64.3)
	N	283	356	372	196	1,207

***** The examiners mark each station based on the candidates performance against the individual objectives associated with the station (such as presenting complaint, approach to the patient etc). A grade between A and E is set for each objective. The grades are then converted to marks (A = 4, B = 3, C = 2, D = 1, E = 0). A score is calculated for each station by multiplying the mark given for each objective by the percentage allocated and then adding them up. So if a candidate were awarded B, C, A, D for four objectives in a station which were weighted 40%, 30%, 20%, 10%, they would score 2.70.

3 x 0.40 = 1.20

2 x 0.30 = 0.60

4 x 0.20 = 0.80

1 x 0.10 = 0.10

Total = 2.70

The examiners also make an overall judgment as to whether the candidate's performance rates as pass, borderline or fail. The overall judgment is used to determine the pass mark for future candidates. The mean scores of previous candidates judged borderline in each station are used to work out the station pass mark. In order to determine if the candidates meet the required standard for the exam, the borderline scores for each of the 14 stations in the exam (pilot stations do not count towards your result) are added up. One standard error of measurement is then added to this score. This creates the total score for the exam. Candidates must meet the required standard of TWO criteria. They must meet or exceed the total score for the exam and achieve the passing score in a minimum of nine stations.

The marks listed in this table refer to scores in the different components of the exam covering communication, examination, history taking and practical skills.

Table 19 **Coefficients for random Effects Logistic Regression Models for First Attempt at CSA for Candidates who have take IELTS and PLAB Part 2 giving unadjusted odds ratios**

(N=1166)	Adjusted OR	95% c.i.	p
BME (compared to White)	2.533	(1.731 ,3.706)	<0.001
EEA (compared to IMG)	0.776	(0.508 ,1.184)	0.239
Constant	0.741	(0.506 ,1.084)	0.123
Intra-cluster Correlation(ICC) for Deanery	0.010		

Table 20 **Coefficient for Random Effects Logistic Regression Models for First Attempt at CSA for Non-UK graduates who have taken IELTS and PLAB adjusted for age, gender, AKT, IELTS and PLAB Part 2**

(N=1166)	Adjusted OR	95% c.i.	p
BME (compared to White)	1.580	(0.878 ,2.845)	0.127
EEA (compared to IMG)	0.968	(0.352 ,2.660)	0.95
Female	0.497	(0.377 ,0.655)	<0.001
Age in Year at Time of CSA Exam	1.101	(1.062 ,1.142)	<0.001
AKT			
Clinical Medicine	0.963	(0.945 ,0.981)	<0.001
Evidence Interpretation	0.996	(0.986 ,1.007)	0.487
Organisational Questions	0.990	(0.977 ,1.002)	0.11
Mark to Pass AKT	1.016	(0.967 ,1.068)	0.522
IELTS			
Reading Score	0.900	(0.731 ,1.108)	0.321
Speaking Score	0.842	(0.673 ,1.054)	0.133
Understanding Score	0.719	(0.590 ,0.877)	0.001
Writing Score	1.004	(0.814 ,1.239)	0.968
PLAB Part 2			
Communication	0.606	(0.440 ,0.836)	0.002
Examination	0.880	(0.654 ,1.185)	0.4
History	0.719	(0.514 ,1.007)	0.055
Practical	0.871	(0.690 ,1.098)	0.241
Constant	632.191	(0.465 ,862069)	0.08
Intra-cluster Correlation(ICC) for Deanery	<0.001		

Table 21 Released from programme (ARCP Outcome 4) by Ethnicity

	White-UK	BME-UK	White-IMG	BME-IMG	White-EEA	BME-EEA	Total
Released	5	11	5	142	4	9	176
(%)	(0.2)	(0.9)	(5.8)	(11.6)	(4.3)	(18.8)	(3.5)
Total	2,484	1,160	86	1,224	93	48	5,095

Table 22 ARCP Outcome 4 by gender

Outcome4	Man	Woman	Total
released	149	27	176
(%)	(7.1)	(0.9)	(3.5)
Total	2,111	2,984	5,095

Table 23 Characteristics of Trainee Released compared to main cohort

	Released from Training			Remainder of Cohort		
	Mean	(sd)	N	Mean	(sd)	N
Age in Years at CSA	38.5	(5.3)	176	31.9	(4.7)	4919
CSA						
Data gathering, Technical and assessment skills	24.4	(3.5)	176	29.5	(3.4)	4919
Clinical management skills	20.8	(3.4)	176	26.7	(3.8)	4919
Interpersonal Skills	21.6	(4.0)	176	29.2	(4.1)	4919
Total	66.8	(9.1)	176	85.4	(10.0)	4919
Total Relative to Pass Mark	-6.5	(9.0)	176	11.9	(9.9)	4919
AKT						
Clin. Med.	71.8	(6.9)	176	77.8	(6.4)	4918
Evid. Inter	63.8	(12.9)	176	76.2	(13.0)	4918
Org. Issues	64.8	(11.2)	176	74.2	(11.6)	4918
Total Relative to Pass Mark	-8.0	(18.0)	176	14.2	(17.5)	4918

Table 24 **Numbers Released by Deanery**

Deanery	Number Released	(%)	N
Defence	0	(0.0)	38
East Midlands	3	(1.0)	292
East of England	9	(2.3)	384
Kent, Surrey and Suss	40	(8.2)	487
London	2	(0.3)	622
Mersey	25	(11.5)	218
Scotland (East)	2	(4.5)	44
Scotland (North)	5	(7.7)	65
Scotland (South East)	2	(2.3)	88
Scotland (West)	9	(4.6)	194
Oxford	3	(1.7)	177
Wessex	1	(0.5)	185
Peninsula	0	(0.0)	106
Severn	0	(0.0)	200
West Midlands	22	(4.0)	549
North Western	23	(5.7)	402
Northern	7	(3.0)	230
Northern Ireland	1	(1.0)	102
Wales	4	(2.1)	191
Yorkshire & Humber	16	(3.6)	440
Not Specified	2	(2.5)	81
Total	176	(3.5)	5,095

Table 25 **Coefficient for Logistic Regression Models of Release from Programme giving unadjusted odds ratios**

(N=5094)	Adjusted OR	95% c.i.	p
BME UK	5.595	(1.917,16.327)	0.002
White IMG	31.627	(8.803,113.631)	<0.001
BME IMG	66.626	(26.919,164.902)	<0.001
White EEA	21.072	(5.497,80.780)	<0.001
BME EEA	119.607	(37.074,385.871)	<0.001
Constant	0.002	(0.001,0.004)	<0.001
ICC	0.166		

Table 26 **Coefficient for Logistic Regression Models of Release from Programme giving estimates of odds ratios by adjusted for gender, age, 1st attempt CSA and AKT**

(N=5094)	Adjusted OR	95% c.i.	p
BME UK	2.828	(0.946 ,8.456)	0.063
White IMG	4.315	(1.031 ,18.061)	0.045
BME IMG	8.316	(3.112 ,22.223)	<0.001
White EEA	6.590	(1.622 ,26.780)	0.008
BME EEA	13.664	(3.646 ,51.207)	<0.001
Female	0.268	(0.169 ,0.425)	<0.001
Age Year Time of CSA Exam	1.030	(0.990 ,1.072)	0.14
CSA			
Data gathering technical & assessment skills	0.953	(0.886 ,1.026)	0.201
Clinical Management skills	0.931	(0.869 ,0.997)	0.042
Interpersonal Skills	0.858	(0.799 ,0.921)	<0.001
AKT			
Clin. Med.	0.980	(0.957 ,1.003)	0.09
Evid. Inter	0.991	(0.977 ,1.004)	0.184
Org. Issues	0.999	(0.983 ,1.015)	0.899
Mark to pass	1.016	(0.943 ,1.095)	0.68
Constant	3.495	(0.000 ,112053.9)	0.813
ICC	0.172		

Table 27 **Outcome of last CSA prior to Release from Programme for Outcome 4 Cases**

	CSA Last Attempt												Total	
	1		2		3		4		5		6			
	Freq	(%)	Freq	(%)	Freq	(%)	Freq	(%)	Freq	(%)	Freq	(%)	Freq	(%)
Fail	2	(15.4)	31	(93.9)	51	(94.4)	33	(84.6)	19	(54.3)	1	(50.0)	137	(77.8)
N (%)	13	(7.4)	33	(18.8)	54	(30.7)	39	(22.2)	35	(19.9)	2	(1.1)	176	(100.0)

Reference List

(1) Wakeford R. International medical graduates' relative under-performance in the MRCGP AKT and CSA examinations. Educ Prim Care 2012; 23(3):148-152.

(2) Tyrer SP, Leung W-C, Smalls J, Katona C. The relationship between medical school of training, age,gender and success in the MRCPsych examinations. Psychiatric Bulletin 2002; 26:257-263.

(3) Dewhurst NG, McManus C, Mollon J, Dacre JE, Vale AJ. 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.

(4) Spike NA, Hays RB. Analysis by training status of performance in the certification examination for Australian family doctors. Med Educ 1999; 33(8):612-615.

(5) Boulet JR, Swanson DB, Cooper RA, Norcini JJ, McKinley DW. A comparison of the characteristics and examination performances of U.S. and non-U.S. citizen international medical graduates who sought Educational Commission for Foreign Medical Graduates certification: 1995-2004. Acad Med 2006; 81(10 Suppl):S116-S119.

(6) The Health and Social care Information Centre- Workforce Directorate. General and Personal Medical Services England 2002-2012. 21-3-0013. The Health and Social Care Information Centre.
Ref Type: Report

(7) Skrondal A, Rabe-Hesketh S. Generalized Latent Variable Modelling: Multilevel, Longitudinal, and Structural Equation Models. Boca Raton, FL: Chapman Hall/CRC; 2004.

(8) Esmail A. Ethnicity and academic performance in the UK. BMJ 2011; 342:d709.

(9) Woolf K, Potts HW, McManus IC. Ethnicity and academic performance in UK trained doctors and medical students: systematic review and meta-analysis. BMJ 2011; 342:d901.

(10) Epstein RM. Assessment in medical education. N Engl J Med 2007; 356(4):387-396.

(11) Esmail A, May C. Commentary: oral exams--get them right or don't bother. BMJ 2000; 320(7231):375.

(12) Patterson F, Denney ML, Wakeford R, Good D. Fair and equal assessment in postgraduate training? A future research agenda. Br J Gen Pract 2011; 61(593):712-713.

Appendix 1: Terms of Reference of the review

Specification for review of the Royal College of General Practitioners (RCGP) Membership of the Royal College of General Practitioners (MRCGP) examination

Introduction

1. The purpose of this document is to outline the specification for a review commissioned by the Registrar of the GMC into the Membership of the Royal College of General Practitioners (MRCGP) examination.

Background

GMC Context

2. The General Medical Council ('the GMC', 'us', 'our', 'we') is the independent regulator for doctors in the UK and the competent authority for awarding qualifications to those who satisfactorily complete training in one of the approved specialties. Our purpose, as set out in Section 1(1)A of the Medical Act 1983, is to protect, promote and maintain the health and safety of the public by ensuring proper standards in the practice of medicine.

3. In order to achieve this a number of key aims and objectives were published in our *Education Strategy 2011-2013* :

   a. Setting and assuring standards, and valuing training.

   b. Promoting effective selection, transition and progression.

   c. Defining outcomes for education and training.

   d. Working with partners and promoting feedback and learning.

4. We have a range of educational standards arising from our statutory duties under Section 34H of the Medical Act 1983 which underpin these aims and objectives, which can be found in the following documentation:-

a. *Tomorrow's Doctors*[vi]

b. *The Trainee Doctor*[vii]

c. *Standards for curricula and assessment systems*[viii]

It is the last of these documents that specifically relates to standards for assessment systems and in particular standards 8, 10 & 12. Looking at these standards together with our responsibilities under the Equalities Act 2010 we are required to be assured that examinations are fair and that they use assessment methodology which is both consistent and the best practice for the type of examination.

5. The GMC approve curricula and supporting assessment blueprints for each of the specialties that it awards a CCT and this approval is against the standards outlined above under 4c.

Postgraduate Training - General Practice

6. The GMC has approved the GP curriculum and assessment blue print against the published standards. The most recent update to this was in 2010[ix,] which included a change to the MRCGP examination.

MRCGP examination

7. The MRCGP is an integrated assessment system, success in which confirms that a doctor has satisfactorily completed specialty training for general practice, and is competent to enter independent practice in the United Kingdom without further supervision. Satisfactory completion of the MRCGP is a compulsory element of the curriculum.

8. The MRCGP comprises three separate components: an Applied Knowledge Test (AKT), a Clinical Skills Assessment (CSA) and Workplace Based Assessment (WPBA), each of which tests different competences using different assessment methods and which together cover the spectrum of knowledge, skills, behaviours and attitudes defined by the GP curriculum.

9. The AKT can be taken during ST2 or ST3, and the CSA can be taken in ST3. The WPBAs are undertaken throughout training, are recorded in the eportfolio and feed into the Annual Review of Competence Progression (ARCP). There is no requirement for success in AKT for eligibility to sit the CSA or for successful assessments to sit either the AKT or CSA, although trainees cannot take the exam

[vi] Tomorrow's Doctors, outcomes and standards for undergraduate medical education September 2009
[vii] The Trainee Doctor Foundation and specialty, including GP training February 2011
[viii] Standards for curricula and assessment systems July 2008 updated April 2010
[ix] GP Curriculum - http://www.gmc-uk.org/education/gp.asp

until the relevant stage (year) of their training, which, for example, would imply they have had to move from ST1 to ST2.

10. As outlined above the MRCGP is a compulsory requirement for success in the CCT curriculum. As such it is a high stakes qualification and if the trainee is not successful he or she will be removed from the training programme (following suitable remediation extension). Given the legal requirements for General Practice, those doctors will also be removed from the local performers list, which means they cannot work in any capacity as a GP. This is not the case for all other specialties where trainees, who leave the training programme without success in the relevant examination, are usually able to continue in a variety of roles within the same specialty. This avoids them becoming deskilled and they can apply via alternative routes to specialist registration, whereas this becomes almost impossible for those in General Practice the longer they are out of the specialty.

11. It has been known for some time that there is a difference in the pass rates for parts of the examination for candidates with different protected characteristics. This is not limited to UK examinations or to General Practice as a specialty[x]. Indeed, a GMC conference in September 2012 – *Being Fair* - included a workshop exploring the impact of place of qualification and ethnicity on progression in UK medical education. The workshop drew on an important academic paper published in the BMJ in 2011 which had concluded that the relationship between ethnicity and academic performance was likely to be complex and multi-factorial. [xi] Among other things, the workshop also noted GMC data showing that – across all postgraduate medical specialties – doctors in training who qualify overseas are twice as likely as their UK counterparts to receive an unsatisfactory outcome in their annual assessments. [xii] The workshop also heard about work the RCGP was doing to investigate this differential outcome in the CSA, and that previous work to look at possible examiner bias had not demonstrated any such effect although it would continue to monitor this.

12. The RCGP publishes statistics on each examination sitting[xiii] and indeed has been ahead of many other specialties in terms of transparency and a willingness to reflect on, and investigate, issues highlighted by such data. The College is due to publish results of a review of "fairness", looking at over 52,000 simulated CSA

---

[x] http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2923.1999.00307.x/full

http://journals.lww.com/academicmedicine/pages/articleviewer.aspx?year=2006&issue=10001&article=00029&type=abstract

http://pb.rcpsych.org/content/26/7/257.full

[xi] *Ethnicity and academic performance in UK trained doctors and medical students: systematic review and meta-analysis, BMJ 2011:342.d901*

[xii] http://www.gmc-uk.org/Being_Fair_report.pdf_50881743.pdf

[xiii] http://www.rcgp.org.uk/gp-training-and-exams/mrcgp-exam-and-assessment-statistics.aspx

consultations. However, allegations of unfairness have persisted, particularly in relation to the CSA. The Registrar of the GMC is therefore commissioning an independent, quantitative, review of recent examination data linked to ARCP data. This data will be provided by the College from eportfolio evidence and will be used to establish the extent of failure rates affecting specific groups of doctors, particularly International Medical Graduates (those with a non UK PMQ) and black and minority ethnic UK and non UK trained doctors. The GMC is undertaking this initial review following specific concerns that have been raised about the MRCGP but it forms part of a planned much wider investigation of the differences in performance at all Postgraduate examinations highlighted by recent research (para 11 above). The reasons for this are likely to be complex and multi factorial and will require more comprehensive data from all colleges and faculties than is currently held by the regulator. The collection of this data is beginning and the research project being formulated for the future.

The Review

13. The review will be of all the AKT and CSA sittings from October 2010 to December 2012. (It was in October 2010 when the process and approach to marking borderline candidates for the MRCGP was changed). It will be important to consider data from 2007 to 2010 to understand the context, but the main focus will be to examine the later data.

Objectives

14. The primary purpose of this review is to identify whether the MRCGP examination fulfils the GMC's standards for assessments. It will also identify any issues relevant to our standards in Domain 3 of *The Trainee Doctor* (Equality, diversity and opportunity) which need to be addressed. (See paragraph 24 below).

15. The review will draw together a definitive data set and identify any patterns in the data. It will also consider the effects of the following:-

- number of sittings of all parts of the examination
- candidate characteristics (gender, age, ethnicity etc)
- candidate PMQ and date from PMQ
- previous assessments of candidates and entry route to registration (PLAB, IELTS and recruitment data if available)
- deanery of training programme
- eportfolio / WPBA outcomes (where available).

16. We envisage that the final report of the review will include:-

a. An overall picture of the examination and its constituent parts;

b. An analysis of candidates who are removed from training, considering the extent to which examination failure contributed to this

c. An analysis of candidates who have had four or more attempts in at least one part of the examination;

d. An analysis of candidates who have failed in the CSA, including an analysis of those who have only failed in the CSA.

e. Advice as to whether the MRCGP examination, and the CSA assessment in particular, meets the GMC's standards.

f. Recommendations for any changes to the MRCGP examination and for any further work in a second phase of research (see also para 12).

17. The completed review will be published.

Deliverables

18. The review should produce a report for the GMC, by 30 June 2013. The report will be used in discussion with key interests with the intention of publication in summer 2013. It will also inform the GMC's planned review of *Standards for curricula and assessment systems*. The report will be shared with key interests before publication to check for factual accuracy.

19. While the report will be used to inform the GMC and other key interests, subject to final sign off by the Registrar, the GMC is happy for it to be published in a peer-review journal, providing this does not prevent the GMC from sharing it with other parties or from disseminating its key findings. Arrangements for this would be discussed and confirmed upon project commissioning. We would not envisage co-authorship with other key interests as it is important for credibility that this piece of work is independent, but organisations that have supplied data or information to the review will be appropriately credited.

20. We expect the project to include:

a. A concise interim report submitted to the GMC by 31st May 2013 which should outline progress, any suggested changes to project scope, and emerging findings.

b. A final report to include but not be limited to:

i. An executive summary.

ii. The aims of the project.

iii. A detailed account of the methods used, including, but not limited to, the methods for searching and analysis and for gathering descriptive and qualitative information.

iv. An evaluative review of the information obtained along with conclusions and recommendations drawn from that analysis.

c. Summary oral presentation of findings to the GMC following completion and sign off of the report.

21. The author of the review will be required to enter into a GMC standard contract before starting the project which will include requirements on data protection and confidentiality including that data analysis can only be undertaken for the purpose of producing the review report (including publishing in a peer-reviewed journal as described in para 19).Once the review has been completed, the reviewer will not have any further access to the data or be authorised to undertake any further analysis.

22. The review will require an analysis of confidential data and the GMC will work with the reviewer, the RCGP and the Postgraduate Deans to discuss the most appropriate way to ensure confidentiality of data. The GMC will review this proposal from a privacy impact perspective to ensure full compliance with the Data Protection Act prior to commencement.

23. The issues covered by the review are many and varied. While this short review should be able to highlight any differences in outcomes between particular groups of doctors, whether such differences raise questions under the mandatory requirements in Domain 3 of *The Trainee Doctor* would probably need further study, including qualitative research, as part of a second phase of work (para 12). However, this quantitative review is expected to identify what those questions might be.

GMC
March 2013

Spike N and Hays RB. Analysis by training status of performance in the certification examination for Australian family doctors. Medical Education, 1999; 33(8):612-5.

Boulet, John R.; Swanson, David B.; Cooper, Richard A.; Norcini, John J.; McKinley, Danette W. Section Editor(s): McIlroy, Jodi PhD; Hemmer, Paul MD. A Comparison of the Characteristics and Examination Performances of U.S. and Non-U.S. Citizen International Medical Graduates who sought Educational Commission for Foreign Medical Graduates Certification: 1995-2004. Academic medicine, 2006; 81(10): S116-S119

S. P. Tyrer, W.-C. Leung, J. Smalls, C. Katona, (2002) The relationship between medical school of training, age, gender and success in the MRCPsych examinations. The Psychiatrist (2002) 26: 257-263